Evolutionary relationships of East African soda lake cichlid fish

Antonia Geraldine Patricia Ford

University College London

A thesis submitted for the degree of Doctor of Philosophy

July 2015

UCL

Department of Genetics, Evolution and Environment

I, Antonia Geraldine Patricia Ford, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Antonia Ford
Candidate

Statement of authorship

Chapter three and part of chapter four have been published in *Molecular Ecology* (Ford AGP, *et al*. 2015) with co-authors Kanchon Dasmahapatra, Lukas Rüber, Karim Gharbi, Timothee Cezard and Julia Day. I wrote the first draft of the manuscript, and comments from all co-authors were incorporated before publication.

I undertook two fieldwork trips to Lakes Natron and Magadi to collect data that formed the basis of the present thesis. Details of the relevant collection permits and local approvals are given in the respective methods sections of the data chapters describing fieldwork.

For the sequencing data that forms chapters three and four, I extracted and conducted quality control tests on the DNA for sequencing. Edinburgh Genomics (sequencing facility) prepared the RAD libraries, and conducted sequencing and preliminary bioinformatics (demultiplexing of Illumina reads, alignment to reference genome using BWA, and archiving of raw reads). I conducted downstream filtering, phylogenomic and population genetic analysis of the data, as well as archiving of the relevant datasets on the Dryad repository. Part of the analysis was conducted at the University of York whilst receiving training from Dr Kanchon Dasmahapatra. Stable isotope analysis was conducted at Scottish Universities Environmental Research Centre (SUERC) Life Sciences Mass Spectrometry Facility. I prepared and measured all samples for analysis, and received training from Dr Jason Newton on the use of the mass spectrometer. All morphological analysis was conducted at UCL and the Natural History Museum, London. I performed all body shape analysis (photographs and landmark digitisation), stomach contents/gut length analysis, and dissections. Pharyngeal jaw photographs for some sites were collected during the course of a Master's research project that I oversaw (Kristoffer Bruun, 2014). Photographs from this project were included in datasets for chapters five and six. I extracted pharyngeal jawbones from additional populations and photographed these, designed the sampling and landmark configuration, and performed the analysis.

“In adaptive radiation and in every part of the whole, wonderful history of life, all the modes and all the factors of evolution are inextricably interwoven. The total process cannot be made simple, but it can be analyzed in part. It is not understood in all its appalling intricacy, but some understanding is in our grasp, and we may trust our own powers to obtain more.”

- Simpson 1953

Abstract

This thesis examines the evolutionary relationships of the *Alcolapia* soda lake cichlid fishes of East Africa. The introduction presents background on the soda lakes in which the cichlids are found, the taxonomy and biology of the fishes, as well as the theoretical background to the study. Chapter two discusses the methods used in the thesis, addressing the benefits and limitations of each, as well as their suitability to the study in hand. Chapter three investigates the phylogenetics and phylogeography of soda lake cichlids sampled at several populations around the soda lakes and a single transplanted population outside of the focal lakes, employing a large genomic dataset generated through restriction site associated DNA (RAD) sequencing, and demonstrates low levels of interspecific genomic differentiation with high levels of ongoing gene flow. Chapter four uses the RAD dataset to test for signals of selection between *Alcolapia* species, employing genome-wide scans and outlier detection to characterise peaks of genomic divergence between species. Chapter five combines morphological (geometric morphometrics) and ecological (stable isotope, stomach contents) data with the RAD dataset from chapter three to consider biologically relevant diversification between *Alcolapia* species, testing for convergence and niche adaptation. Chapter six examines the ecomorphology of the soda lake fishes at an intraspecific level, testing for effects of geography and environment on morphological differentiation between populations. Finally, chapter seven draws together the conclusions inferred from the thesis, and discusses possible future directions for research in this system.

TABLE OF CONTENTS

CHAPTER THREE –

LIST OF FIGURES

LIST OF TABLES

Acknowledgements

My grateful thanks first and foremost to my supervisors, Julia Day and Lukas Rüber who have provided unstinting guidance, support and advice throughout my doctoral research. Thank you also to Kanchon Dasmahapatra, who provided invaluable training in analysis of next generation sequence data, as well as useful advice on the general scientific process and career development. Thank you to Kanchon and Ruth for hosting me on visits to York. Members of the Day lab and UCL Department of GEE provided help and friendship along the way: Siobhan Cox, Katherine Brown, Claire Peart, Simon Maddock, Jake Morris, Adam Britton, Marcela Randau, Thomas Halliday, Andrew Cuff, Fabian Zimmer, and Kate Brown.

Several collaborators were instrumental to this research project. John Dominic Balarin provided useful planning advice for fieldwork, as well as kind hospitality in Dar Es Salaam, and unfailing enthusiasm for tilapia research. The Tanzania Fisheries Research Institute, COSTECH, National Museums of Kenya and NCST are acknowledged for research permits and facilitating fieldwork. Ben Ngatunga of TAFIRI and Marc Baker of Carbon Tanzania were generous in sharing their time and advice for working on the soda lakes. Edward Njagi and Dorothy Nyingi of the Nairobi Museum were of great help in planning and conducting the Kenyan part of the fieldwork. Jason Newton at SUERC provided valuable advice and training in stable isotope analysis. Edinburgh Genomics provided useful sequencing assistance, particularly Cathlene Eland in the preparation of DNA samples, and Timothee Cezard in bioinformatics analysis. Thank you to Mark Wilkinson and David Gower at the NHM for kind use of laboratory facilities and equipment for morphometric analyses. Thanks also to Janet Hope and Ian Patmore in the UCL Geography department for advice and loan of equipment. A general thank you to the wider cichlid research community, who have proved supportive and interested in this research at several conferences and symposia. Several software program authors were also kind in offering advice on the correct use of their respective programs: Olivier Francois, Oscar Gaggiotti, David Polly and Pablo Cingolani.

Funding is acknowledged from the UCL graduate school, UCL department of GEE, the Systematics Research Fund, the Genetics Society (fieldwork, training, and conference grants), the British Ecological Society (travel grant), and a NERC grant-in-kind for use of the Life Sciences Mass Spectrometry Facility at East Kilbride.

Finally, thank you friends and family who have been supportive and understanding over the last four years. Special thanks to Chris, Kingsley, Millie, Minty, and my parents.

This thesis is dedicated to my parents, Richard and Susan Ford,
who have always believed in me.

Chapter one

Introduction

Adaptive radiation

The comparatively high diversity displayed by groups of taxa in certain environments has long drawn interest from an evolutionary perspective. The components of adaptive radiation, ecological diversity in a diverging lineage, were noted in Darwin's discussion of natural selection and oceanic islands:

"…the simple circumstance that the more diversified the descendants from any one species become in structure, constitution, and habits, by so much, will they be better enabled to seize on many and widely diversified places in the polity of nature, and so be enabled to increase in numbers." (Darwin 1859)

The term adaptive radiation was introduced by (Osborn 1902), however it was some 50 years later that the term was formally defined by Simpson (1953), and the key definition of rapid divergence was introduced:

"So far as adaptive radiation can be distinguished from progressive occupation of numerous zones, a phenomenon with which it intergrades, the distinction is that adaptive radiation strictly speaking refers to more or less simultaneous divergence of numerous lines all from much the same ancestral adaptive type into different also diverging adaptive zones" (Simpson 1953, p 223).

More recently defined as the evolution of ecological diversity in a rapidly multiplying lineage (Schluter 2000), cases of adaptive radiation are increasingly of interest for the study of speciation, and genetic and ecological structure of species diversity (Hudson *et al.* 2011). Adaptive radiation comprises both speciation and the adaptation of those resultant species to a diversity of ecological niches (Gavrilets & Losos 2009). Notable examples of such radiations include the Hawaiian silversword alliance (Robichaux *et al.* 1990), the highly speciose haplochromine cichlid fishes of the East African Great Lakes (Nagl *et al.* 2000; Salzburger *et al.* 2005) and, perhaps most famously, Darwin's finches (Grant & Grant 2008), Figure 1.1.

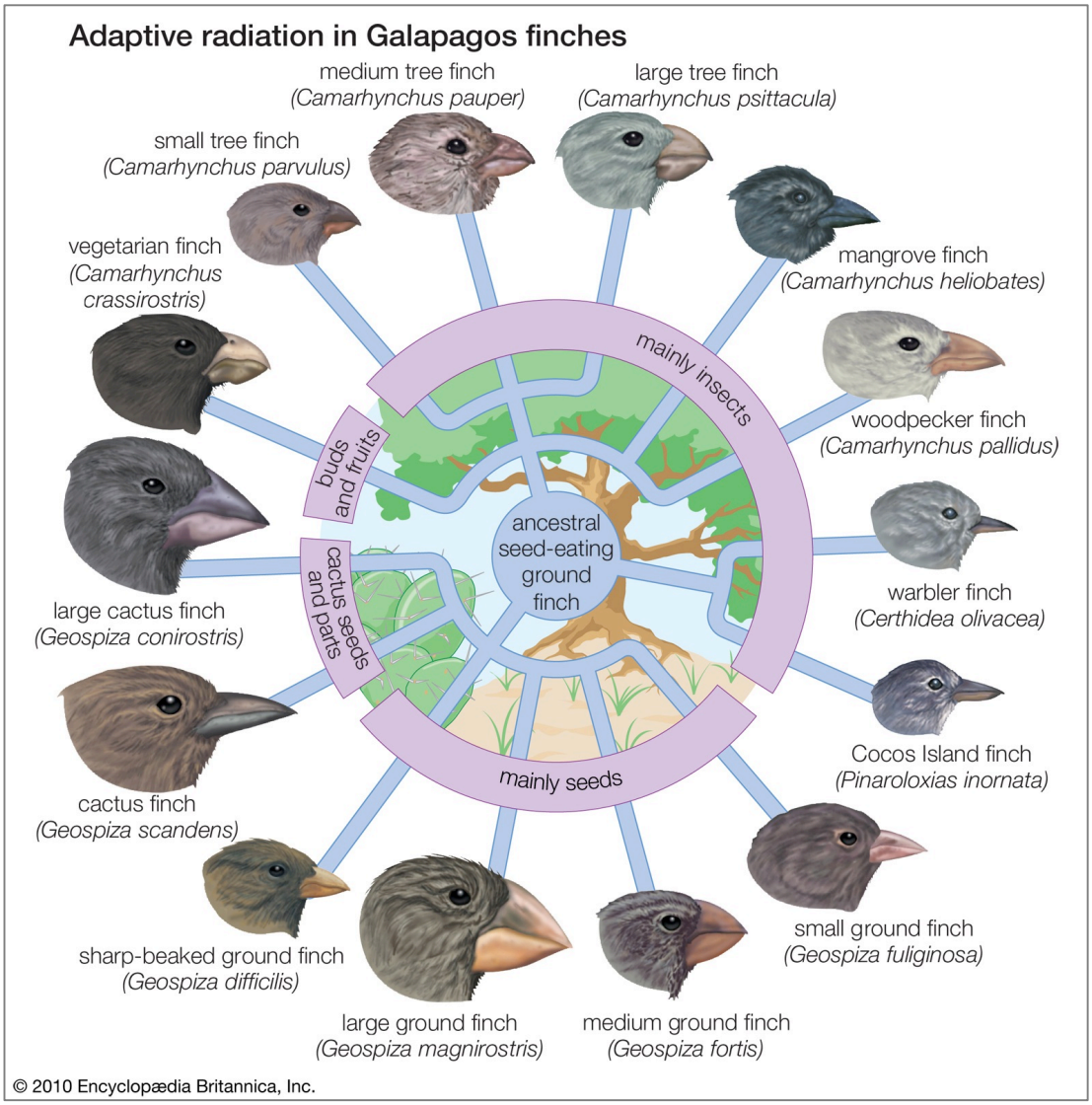


Figure 1.1. Adaptive radiation in Darwin's finches.

Fourteen species of Galapagos finches that evolved from a common ancestor, demonstrating adaptive divergence through phenotype-environment correlation of bill shape to resource type. Reproduced in accordance with the copyright license of Encyclopædia Britannica, Inc.

The process of such radiation includes speciation and phenotypic adaptation to different environments. Three processes are suggested to be responsible: divergent selection; evolution of reproductive isolation between populations exploiting different resources; and ecological character displacement, where competition between similar phenotypes for food drives divergence to exploit new resources (Schluter 1993). Several criteria have subsequently been proposed as critical test of adaptive radiation: common ancestry of component species; phenotype-environment correlation; trait utility, where traits exhibit performance or fitness advantages in the

relevant environment; and rapid speciation, with bursts of species formation contemporaneous to ecological and phenotypic divergence (Schluter 2000).

As ecological character displacement will be aided by lack of competing taxa (Schluter 1993), the theory of adaptive radiation often considers positions of ecological opportunity, where the ancestral species occurs in regions of resource under-utilisation such as colonisation of unpopulated areas, extinction of other groups, or evolution of a 'key innovation'. Hence, island systems (or island-like isolated regions such as constrained water bodies), which may see rapid diversification of early colonisers, have been the focus of much research and include some of the most notable diversity of species radiations. Attempts have been made to identify the influences and properties of such adaptive radiations (Gavrilets & Losos 2009; Birand *et al.* 2012), to explain the lack of diversity in 'failed radiations' (Seehausen 2006; Nosil *et al.* 2009), and to characterise species diversity in non-adaptive radiations that show no ecological diversification (e.g., Wellenreuther & Sánchez-Guillén 2015). These studies have resulted in a suite of requirements being proposed to enable or predispose an adaptive radiation, however most researchers caution that each situation needs to be considered on a case-by-case basis and the role of stochasticity must be consdiered (e.g., Seehausen 2007). The timeline for ecological divergence may vary between radiations, and differs substantially in some of the most notable cases of the phenomenon (Table 1.1).

Table 1.1 Notable adaptive radiations.

Group	Timescale	Species	Primary axis of divergence
Darwin's finches	1.5 my (Petren et al. 2005)	14	Bill shape (resource use)
Hawaiian honeycreepers	2-4 my (Lerner et al. 2011)	~50	Bill shape (resource use)
Hawaiian silverwords	5.2 my (Baldwin et al. 1998)	28	Leaf shape / structure
Heliconius butterfiies	11-13 my (Kozak et al. 2015)	46	Wing colour patterns
Caribbean *Anolis* lizards	<40 my (Losos et al. 2004)	~140	Tree trunk-ground
Threespine stickleback	<40 ky (Mckinnon et al. 2004)	~1000 pairs	Anadromous-freshwater
Arctic Charr, Thingvallavatn	<10 ky (Snorrason et al. 2004)	4 morphs	Benthic-limnetic
Antarctic notothenioids	16-30 my (Matschiner et al . 2011)	~120	Buoyancy adaptations
Lake Victoria region haplochromine cichlids	<250 ky (Verheyen et al. 2003)	~500	Male colour, trophic diversification
Lake Malawi haplochromine cichlids	1-5 my (Koblmüller et al. 2008)	>600	Male colour, trophic diversification
Lake Tanganyika cichlids	9-12 my (Salzburger et al. 2005)	~200	Male colour, trophic diversification

Although Schluter's (2000) definition of adaptive radiation included speciation at the same time as ecological and phenotypic divergence, speciation may not necessarily

be complete at the time of radiation, but ecological divergence may simply be driving the evolution of reproductive isolation (Nosil 2012). It has been suggested that in some cases of adaptive radiation speciation may never occur: proposed as an alternative to ecological speciation, where adaptive radiation is an intraspecific phenomenon in which speciation happens well after adaptation if at all (Givnish 1997; Bezault *et al.* 2011). Certainly, several cases of recently diverged lineages with morphologically defined species or morphs are considered adaptive radiations despite low genomic differentiation and ongoing gene flow (Parchman *et al.* 2006; Samonte *et al.* 2007; Elmer *et al.* 2010b; Martin & Feinstein 2014; Kahilainen *et al.* 2014; Ford *et al.* 2015).

Ecological speciation may occur in sympatry or allopatry, although it seems that such speciation in extreme cases of either (pure sympatry or allopatry) are rare, and that ecological speciation more frequently occurs on a gradient of separation with some degree of population disconnection followed by secondary contact (Schluter 2001). Even in cases of presumed sympatry in entirely overlapping ranges, there may not be complete syntopy (Rivas 1964), and factors of microallopatry (habitat subdivision) may be at play (Rico & Turner 2002; Habel *et al.* 2012). However, the definition of sympatry will vary between viewpoints from biogeography (overlapping ranges) and population genetics (likelihood of meeting and mating with other individuals regardless of geographic separation) (Fitzpatrick *et al.* 2008).

Cases of sympatric speciation are hard to demonstrate definitively, given the difficulty of ruling out some separation in the time course of divergence (Coyne & Orr 2004). However, even cases with a large degree of support require theoretical explanation for how speciation may occur in the face of homogenising gene flow (Fitzpatrick *et al.* 2008). Divergence hitchhiking has also been posited to explain how regions of the genome may resist the effects of recombination from between-population mixing (Via 2011). Furthermore, rapid ecological speciation may rely on existing genetic variation (Barrett & Schluter 2008), and a recent large-scale genomic study suggests this may the case in East African haplo-tilapiine cichlid fish, with accumulation of genetic variation under relaxed constraint preceding radiation and divergent selection acting on many genes (Brawand *et al.* 2014).

Adaptive radiation is often characterised by interspecific hybridisation after the onset of speciation (Grant & Grant 2008; Lamichhaney *et al.* 2015) and the hybrid swarm theory of adaptive radiation suggests that in the colonisation of new environments, hybridisation increases response to selection and diversification under divergent selection – so may in fact promote the formation of new species (Seehausen 2004).

Schluter (2000) cautioned that the presence of diversity in different environments does not necessarily indicate an adaptive radiation, giving the example of adaptation to local environment in damselfly larvae of 73 species where morphology differs between lake types but species are not ecologically differentiated within the same lake environment. Furthermore, a recent review of the development of adaptive radiation research has cautioned the conflation of radiation with explosive diversification (Givnish 2015). A recent scientometric investigation of adaptive radiation studies highlighted the fluidity of the adaptive radiation concept and lack of clear threshold criteria (Soulebeau *et al.* 2015). The authors emphasised the lack of rigorous testing of the adaptive radiation hypothesis in most studies and criticised the high proportion of studies not including phylogenetic analysis of any kind (Soulebeau *et al.* 2015). While it may be useful to ensure cases of diversification meet minimum criteria before describing them as adaptive radiations, it is still undoubtedly of interest to describe and characterise the patterns and processes driving such diversification. Certainly, the increasing use of and methods applied to adaptive radiations to test morphological divergence e.g., (Harmon *et al.* 2003) and the advent of large genomic datasets are allowing diversification to be more rigorously investigated than previously possible.

Cichlid speciation and adaptation

The species abundance and morphological diversity of cichlid radiations of the African Great Lakes have long been the focus of evolutionary study (Fryer & Iles 1969; 1972; Greenwood 1974; Van Couvering 1982). The incredibly rapid speciation process has drawn great interest, based on the high species number and short timeframe compared to the other famed adaptive radiations (Table 1.1). Cichlidae comprise one of the most species-rich family of vertebrates, with 1,700 valid species currently described (Eschmeyer 2015) but with estimates ranging up to 3,000 (discussed in Salzburger & Meyer 2004). Explanations proffered for cichlid propensity to speciate have included: the pharyngeal jaw apparatus that uncouples resource foraging from processing (Liem 1973); a syngameon phase of diversification (Seehausen 2004) and hybridisation (Loh *et al.* 2013); enriched adaptive variation removing genetic constraints (Seehausen 2006); polygenic basis for ecological trait and mate preference (Dieckmann & Doebeli 1999; Kondrashov & Kondrashov 1999; Kocher 2004); asymmetric parental investment (maternal mouth brooding) leading to strong sexual selection (Kocher 2004); genetic conflict via variable sex-determining mechanisms (Lande *et al.* 2001); exploitation of novel

habitats and ecological opportunity (Hulsey *et al.* 2010); and phenotypic plasticity (Stauffer & Van Snik Gray 2004).

Although early study focused on the African Great Lakes, cichlids have radiated in many other locations (Table 1.2) and current studies on cichlid speciation increasingly focus on smaller isolated lakes containing young and geographically restricted radiations (e.g., Schliewen *et al.* 2001; Elmer *et al.* 2010a; b; Recknagel *et al.* 2013a). Recently diverged species offer an ideal testing ground to study the initial stages of speciation (Hudson *et al.* 2011), since the stages at which alternative speciation mechanisms are relevant are easier to trace than in more advanced radiations, making it possible to elucidate the mechanisms of adaptations to new environmental settings. Furthermore, several of these radiations offer parallel conditions, where an ancestor independently invaded multiple locations with the same environments and diverged along similar axes, such as the crater lake cichlids of Nicaragua (Kautt *et al.* 2012; Elmer *et al.* 2014), Uganda (Machado-Schiaffino *et al.* 2015) and Cameroon (Schliewen *et al.* 1994). Such replicate adaptive radiations, where similar morphotypes evolve in response to similar environments, provide evidence of ecological character displacement and negate the alternative of speciation independent of divergent selection (Schluter & Mcphail 1993).

Table 1.2. Lacustrine cichlid radiations outside the African Great Lakes.

Lake	Type	Size (km^2)	Species
Barombi Mbo, Cameroon	Crater lake	4.15	11 (Schliewen *et al.* 1994)
Ejagaham, Cameroon	Crater lake	0.5	2 and 3* (Martin *et al.* 2015)
Bermin, Cameroon	Crater lake	0.6	9 (Schliewen *et al.* 1994)
Mweru, Zambia/DRC	Freshwater	5,120	13 morphs (Stelkens & Seehausen 2009)
Nabugabo, Uganda	Freshwater	220	7 (Bezault *et al*. 2011)
Natron, Tanzania	Soda lake	81–804†	3 (Seegers & Tichy 1999)
Apoyo, Nicaragua	Crater lake	21	6 (Geiger *et al*. 2010)
Xiloá, Nicaragua	Crater lake	8	4 (Recknagel *et al*. 2013b)

*denotes separate radiations.

†Lake area is highly variable and dependent on rains.

Of the species-rich cichlid species flocks from the African Great Lakes, almost all species are endemic to a single lake. Parallel evolution of similar body morphologies in each lake has been attributed to convergent evolution in response to similar environmental pressures exhibiting divergent selection (Kocher *et al.* 1993), Figure 1.2.

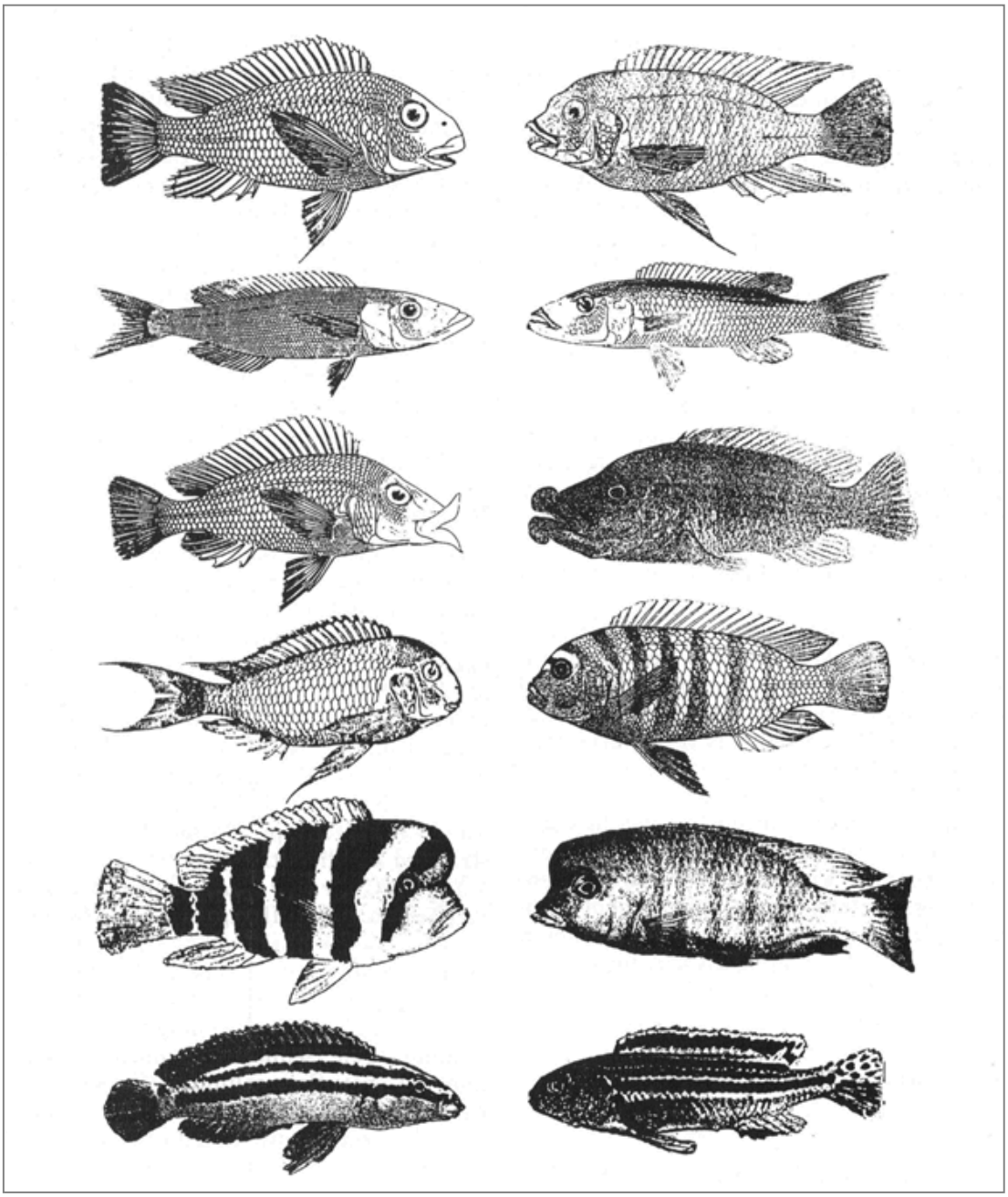


Figure 1.2 Convergent evolution of African cichlid fish.

Reproduced from Kocher *et al.* 1993 with permission from Elsevier. Tanganyika (left) and Malawi (right) cichlid pairs display convergent phenotypes of (top to bottom): rasping jaw, fusiform body, fleshy lips, mbuna specialisation, nuchal hump, horizontal stripes.

The large haplochromine species flocks arose in Lake Tanganyika and later colonised Lakes Malawi and Victoria (Salzburger *et al.* 2005). It was thought that, given the lack of shared species between the three lakes, each of the flocks was monophyletic and shared phenotypes that arose solely through convergence. However phylogenetic analysis has suggested that the lakes are not isolated entities, with shared genetic polymorphisms between species of all three lakes (Loh *et al.* 2013). Furthermore, a recent colonisation of Lake Victoria by a Lake Tanganyika species suggests that the hydrology between East African lakes is permeable to species transfer (Meyer *et al.* 2015). Recent developments in genomic analysis have also revealed sharing of parts of the genome between lakes (Brawand *et al.* 2014). The advent of next generation sequencing has provided a fine level of detail to consider the genetic patterns of radiation, allowing for improved phylogenetic resolution (Wagner *et al.* 2013), identification of regions under divergent selection between morphs (Fan *et al.* 2011), and examination of genomic mechanisms underlying phenotypic evolution (Brawand *et al.* 2014).

Riverine cichlids are substantially less diverse than those of the Great Lakes, possibly due to restricted ecological opportunity afforded in river habitats (Schluter 2000; Seehausen 2007) and because radiation size scales with area (Wagner *et al.* 2014). It has also recently been suggested that the greater temporal instability and reduced spatial dimensionality of rivers may make ecological speciation difficult compared to lacustrine environments (Seehausen 2015). In particular, the *Oreochromis* are largely riverine, and although present in the African Great Lakes have only speciated in two lakes, Lakes Albert and Malawi, in which they co-occur with haplochromine cichlids (Seehausen 2007). The only other instance where they appear to have radiated and show patterns of trophic divergence and male colouration typical of the Great Lake species flocks, is in a lake where there are no other genera present: a small radiation of cichlids (previously considered a subgenus of *Oreochromis*) in the Natron-Magadi basin, the *Alcolapia*.

Natron-Magadi basin

The *Alcolapia* cichlid genus is endemic to the springs and lagoons of Lakes Natron and Magadi, thriving in the hostile environment of hydrothermal saline-alkaline waters.

Geology and fossils

Lakes Natron and Magadi form a single North-South basin in Eastern Africa, across the Kenya-Tanzania border and were previously a single palaeolake, Orolonga (Figure 1.3). The basin formed 1.7 million years ago (MYA), and the palaeolake formed ~700 KYA (Eugster 1986). The maximum depth of freshwater palaeolake Orolonga is estimated to have been 50-60m (Roberts *et al.* 1993). Lake Magadi and Lake Natron reached their highest level 10-13 KYA (Roberts *et al.* 1993; Williamson *et al.* 1993), but shortly afterwards water levels dropped below the 635m altitude barrier and the lakes were separated during an arid event in the equatorial extension of the Younger Dryas event (Williamson *et al.* 1993; Behr 2002). The hypersaline conditions of the current lakes are thought to arisen 7-9 KYA as water levels contracted (Butzer *et al.* 1972; Roberts *et al.* 1993).

It appears that the palaeolake was previously occupied by cichlids, as cichlid fossils with similar morphology but larger size than *Alcolapia* specimens (suggesting life in freshwater conditions) have been found in deposits surrounding Lake Magadi. Fossils were found in the High Magadi Beds, about 12m above the present lake level, in a 5-inch (12 cm) band of clay layer over lava (Coe 1966). The fossil fish were initially described by the Kenya Coryndon Museum fish warden (H Copley) as *Tilapia* (=*Oreochromis) nilotica* (White 1953; Coe 1966), though he later noted that the fossils were *Tilapia* (=*Alcolapia*) *grahami,* but considered *T. grahami* to be a relict, stunted, form of *T. nilotica* (Copley 1958). Other workers examining the fossils concluded that they exhibited general features of modern day *A. grahami* (Whitehead, pers comms in: Coe 1966). The 8-inch (~200mm) fossils are somewhat larger than modern day *Alcolapia* (maximum SL: 100mm), suggesting they may have inhabited freshwater, and ^{14}C dating is in agreement, dating the fossils at 9,120 (±170) years old, at which time the lake waters are thought to have been substantially less alkaline than the present conditions (White 1953; Butzer *et al.* 1972; Roberts *et al.* 1993). The current location of these fossils is not known, although they do not appear to be catalogued at the Natural History Museum (London), despite having been examined by the Head of Ichthyology on their discovery (Coe 1966). Later researchers found fossils in Magadi beds on the southeast and northwest shores of Lake Magadi, which were also dated to ±10,000 years (^{14}C), but had a smaller length (SL: 30-100mm) (Tichy & Seegers 1999). Examination of fossil scales and otoliths from these later collections suggested a single morphotype identical to *Alcolapia*, and similar to *O. mossambicus* (Tichy & Seegers 1999). There are also fossiliferous beds at the same height bordering Lake Natron (Coe 1969; Trewavas 1983) but no fossils have been found (Tichy &

Seegers 1999), although to my knowledge no detailed surveys have been conducted in the region.

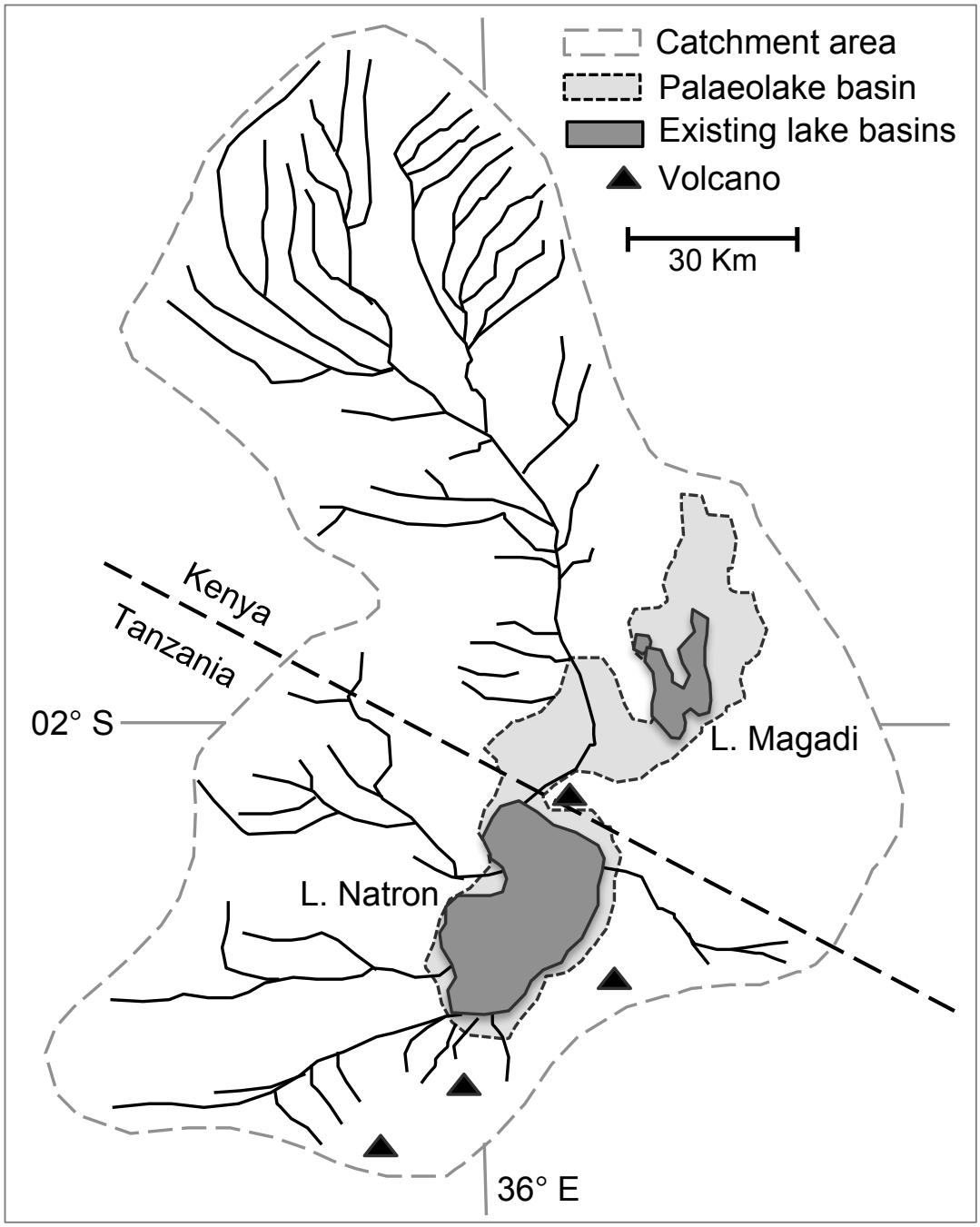


Figure 1.3. Catchment area of the Natron-Magadi basin.

Hydrological system and extent of palaeolake and modern lake boundaries.

Modified from Roberts *et al.* (1993).

Hydrochemistry

While all the African Great Lakes and rift lakes tend to be alkaline, with a pH range of 7-9 (reviewed in Salzburger *et al.* 2014), Lakes Natron and Magadi are exceptional, with highly alkaline and saline waters. Both the lakes are shallow endorheic basins, with Lake Natron having an average lake area of 398 km^2 varying from 81–804 km^2 (Tebbs *et al.* 2013), and Lake Magadi ranging from 75–108 km^2 in dry to wet seasons (Jones *et al.* 1977; Vanden Bossche & Bernacsek 1990). The lakes are subject to substantial climatic effects, with a negative evaporative balance (Burrough & Thomas 2009). The combined Natron-Magadi basin has a negative disequilibrium balance of -28, indicating the discrepancy between present hydrological conditions and the climatic regimen required to maintain full-lake conditions (Hoy & Stephens 1979; Burrough & Thomas 2009). The area is volcanic with alkaline hydrothermal springs containing high levels of salts and precipitates feeding into the lagoons (Williamson *et al.* 1993). A thick layer of crystalline trona (sodium sesquicarbonate precipitate) covers most of the lakes' surfaces, forming a solid covering separating lagoons of permanent open water close to the shore (Kaufman *et al.* 1990).

The trona deposit has commercial value, and is processed by a soda plant at Lake Magadi. The Magadi soda plant has been in place since 1911 and began manufacturing in 1913 (McKetta 1995), and since 2005 has been owned and managed by Tata chemicals, which produces both soda ash (refined, decomposed sodium sesquicarbonate) and common salt (sodium chloride) (Tata Chemicals Ltd 2015). It is the second largest producer of trona in the world (Maxon & Ofcansky 2014). The Magadi township located on the Eastern shore of the lake has a population of ~5000 (Tata Chemicals Ltd 2011).

As well as being considerably smaller than Lake Natron (covering only ~20% of the area), Lake Magadi also differs in having no perennial inflowing streams, while Lake Natron has two inflowing rivers, Peninj and Ewaso Ngiro, as well as several perennial streams (Olaka *et al.* 2010). This factor not only has implications for hydrochemical variability between the two lakes, but also in terms of niche space available, as inflowing rivers and streams provide longer stretches of continuous open water than the volcanic springs. The influx into the Natron system is variable between seasons, especially as the main inflowing river Ewaso Ngiro periodically dries up completely, but the typical hydrological contributions are as follows: direct rainfall onto lake surface: 39%; Ewaso Ngiro river: 27%; saline springs: 25%; other streams: 9% (Tebbs *et al.* 2013).

The conditions in the volcanic springs represent one of the most hostile environments to support fish life, with variables measured for the present study including water temperatures of 30–42.8°C, pH 9-11.5, fluctuating dissolved oxygen (DO) levels of 0.08-6.46mg/L, and high salt concentrations (>20 ppt). These salt concentrations represent approximately half the osmolarity of seawater levels (Lykkeboe *et al.* 1975). Conditions vary geographically between springs and lagoons (Wilson *et al.* 2004; present study), but also diurnally and seasonally with temperature and oxygen varying with levels of sunlight (Coe 1966; 1969).

Natron-Magadi biodiversity

Alcolapia are the only fish species occurring in the lakes and springs of the Natron-Magadi basin, meaning there is no competition or predation from other fish species, and no potential for hybridisation with other cichlids. The southern Ewaso Ngiro river contains *Barbus* and *Clarias* species (Seegers *et al.* 2003). Reports of *Oreochromis spilurus* occurring in the southern Ewaso Ngiro (Fischer 1883; Trewavas 1983), do not appear to be substantiated (Seegers *et al.* 2003).

The basins host a variety of birdlife, most notably flamingos, with Lake Natron being the only successful breeding site of the lesser flamingo, *Phoeniconaias minor*, and recognised as a RAMSAR wetland of International Importance (Tebbs *et al.* 2013). The lake supports a population of 2-4 million lesser flamingos (75% of the global population), and around 100,000 individuals of other species including white stork, ibises, African spoonbill, stilts, plovers, lapwings, sandpipers, and terns (Soussa *et al.* 2010).

Primary productivity rates are extremely high in soda lakes due to high temperature, light intensity and CO_2 levels (Grant 2006). The pH levels of the water preclude the presence of vascular plants, and the only eukaryotic phototrophs present are unicellular green algae, with the bulk of production by cyanobacteria (Mikhodyuk *et al.* 2008b), which give the lakes a characteristic red colour and are the main food source of the flamingos (Grant 2006; Kadigi *et al.* 2012). Several endemic bacterial and microbial strains have been isolated from the lakes (e.g., (Jones *et al.* 1998; Mikhodyuk *et al.* 2008a; b; Muruga 2013; Muruga & Anyango 2013; Zavarzina *et al.* 2013; Muruga *et al.* 2014).

The basin wildlife is currently threatened by plans of soda mine plant development at Lake Natron, and construction of a hydroelectric dam on the Southern Ewaso Ngiro River for power and irrigation (Soussa *et al.* 2010; Kadigi *et al.* 2012; Tebbs *et al.* 2013).

The *Alcolapia* species flock

Currently the *Alcolapia* cichlid radiation includes four described species, with three sympatric species in Lake Natron (*A. alcalica, A. latilabris, A. ndalalani*) and a further species (*A. grahami*) restricted to Lake Magadi. However, there is considerable phenotypic variability across the group, including in the morphology of oral and pharyngeal jaws (Tichy & Seegers 1999), as well as variation in male colour both between species and between intraspecific morphs and populations (Coe 1969; Seegers & Tichy 1999; Figure 1.4).

Figure 1.4. Morphological and colour diversity of the *Alcolapia* radiation.
(Photographs copyright JJ Day, L Rüber, AGP Ford). Clockwise from top right: male and female *A. grahami*; male *A. alcalica* yellow morph from Lake Natron western shore; male *A. alcalica* blue morph from Lake Natron eastern shore; *A. alcalica* trophic (upturned mouth) morph from Lake Natron eastern shore; *A. alcalica* dark morph; *A. latilabris* dark morph; *A. latilabris* from Lake Natron southern lagoon; *A. ndalalani* from Lake Natron southern lagoon.

Alcolapia occur along the springs feeding into the lakes, although they also inhabit edges of the lagoons where the springwater meets the lake body (Narahara *et al.* 1996; Seegers & Tichy 1999). The lagoons are intermittently connected during heavy floods in the rainy season, which may allow migration of *Alcolapia* between populations usually restricted to isolated lagoons during the dry season (Seegers & Tichy 1999; Zaccara *et al.* 2014).

Taxonomy and species descriptions

The *Alcolapia* soda lake cichlids were first named in the literature by F. Hilgendorf in 1905, who described *A. alcalica* originally as *Tilapia alcalica* (Arabic; "alkaline") based on a population found by O. Neumann in 1893 in the volcanic springs at Ndalalani on the south-west shore of lake Natron (Hilgendorf 1905). The type material on which Hilgendorf based the description is accessioned at the Natural History Museum, London (Syntypes: BMNH 1905.7.25.29-31) and the Natural History Museum, Berlin (Lectotype: ZMB 16337; paralectotypes: ZMB 31.905). However, specimens had previously been collected from hot springs in the Natron basin and described as *Chromis niloticus mossambicus* (Fischer 1883), which are now thought to have been *A. alcolapia*, (Trewavas 1983) although the type material collected has since been lost (discussed in Seegers & Tichy 1999). *Alcolapia grahami* was described shortly afterwards, discovered by F. W. Graham in Magadi, and originally thought to be *Tilapia mossambicus* (Woodhouse 1912), but on further examination described as a new species and named after the discoverer, as *Tilapia grahami* (Boulenger 1912a). Thys van den Audenaerde (1968) later described the two species as subgenera within the *Tilapia* genus, erecting the subgenus *Alcolapia*, also including a third species *Tilapia amphimelas* within this subgenus. Trewavas (1973) further subdivided the genus *Tilapia*, keeping substrate brooders within the genus *Tilapia* but placing all mouthbrooders in the newly elevated genus *Sarotherodon* (previously subgenus, Rüppell 1852) specifically including *Sarotherodon amphimelas* and *Sarotherodon alcalicus* (Trewavas 1973). Trewavas (1982) later amended this classification, moving the maternal mouth brooders to *Oreochromis* (Günther 1889), and retaining the male and biparental mouth brooders within *Sarotherodon* (Trewavas 1982a). At this time, Trewavas kept Thys van den Audenaerde's (1968) subgenus *Alcolapia* for *O. grahami* and *O. alcalicus*, but designated them as subspecies: *Oreochromis (Alcolapia) alcalicus alcalicus* and *Oreochromis (Alcolapia) alcalicus grahami.* This designation was based on the morphological similarity of the species types: "Structurally these forms are so similar

that on the basis of preserved specimens they would be treated as one taxon." (Trewavas 1983). Additionally, she erected the new subgenus *Vallicola* (Latin; "valley-dweller") for *O. (V.) amphimelas* (Trewavas 1983).

From its discovery until the end of the twentieth century, the *Alcolapia* flock received little attention in terms of populations and geography, with morphological and behavioural descriptions all based on collections from limited localities typically on the eastern shore (with the exception of the type locality in the south). Seegers and Tichy (1999) conducted the first detailed survey of the Natron-Magadi basin, collecting at several localities and describing additional morphs of *O. a. alcalicus* and variability within *O. a. grahami* populations, as well as describing two new species. These descriptions retained Trewavas' (1983) taxonomy of *Alcolapia* as a subgenus within *Oreochromis*, but elevated the constituent taxa to species (rather than subspecies) status. The two new species descriptions were based on morphology and colour, being *O. a. ndalalani* (named after the type locality; "area with two streams" in Masaai), and *O. a. latilabris* (Latin; "wide-lip), including description of two separate morphs (dark and light form) of the latter species. This placement was subsequently revised the same year, following molecular analysis of mtDNA (partial D-Loop and cytochrome b), showing substantial molecular divergence of the *Alcolapia* species from other *Oreochromis* species, and the subgenus *Alcolapia* was raised to genus rank (Seegers *et al.* 1999). At the time, the Natron species was designated *Alcolapia alcalicus* (retaining the masculine agreement from *Oreochromis alcalicus*), but this was later corrected to *Alcolapia alcalica* (Seegers 2008). The taxonomy of Seegers *et al.* (1999) (with the later 2008 correction) represents the valid taxonomic names and species status in use today (Eschmeyer 2015). The species defining features and colouration are given in Table 1.3, and radiographs (reproduced) of the type material in Figure 1.5.

Table 1.3. *Alcolapia* species defining features (Seegers and Tichy 1999).

	Description	Colouration
A. alcalica	Head usually conical Mouth terminal to subterminal, snout retrognathous Head narrower than in *A. ndalalani* and *A. latilabris*, but lower jaw longer. Teeth not visible when mouth is closed	Males: breast and vent bluish to yellow or golden yellow.
A. grahami	Head conical Mouth terminal, snout prognathous Lips wide but flat Lower jaw more rounded and rostrally turned upwards. Teeth not visible when mouth is closed	Generally grey-mauve with white to pink breast and belly. Dark vertical bars wider than in Natron species.
A. latilabris	Unique head morphology Head long and wide Snout remarkably long and wide Mouth subterminal to inferior with broad jaws and short lower jaw Pronounced concavity of chin region Teeth densely set and visible even when mouth closed	Males: dark olive to light yellow on body. Belly, throat and cheek lighter, flank light olive bluish-green or reddish, 8-10 dark olive vertical bars on posterior part of body.
A. ndalalani	Head short Mouth subterminal Lower jaw shorter than *A. alcalica* but longer than in *A. latilabris* Eye large More dorsal spines than *A. alcalica* and *A. latilabris*	Males: intense orange on cheeks, throat, breast, belly, lower flanks. (bright red in dominant males)

Figure 1.5. Radiograph of *Alcolapia* type material. Reproduced from Seegers and Tichy (1999) with permission from Pfeil-Verlag. A) *A. alcalica* lectotype (ZMB 16.377); B) *A. grahami* lectotype (BMNH 1911.1.21); C) *A. ndalalani* holotype (ZMB 32.883); D) *A. latilabris* holotype (ZMB 32.888).

Alcolapia physiology and adaptation

The hostile nature of the soda lakes has drawn interest on the extremes that *Alcolapia* are able to tolerate and, more recently, investigation of the adaptations that allow them to survive in such an environment. Most focus has entirely centred on *A. grahami*, with only very limited study of *A. alcalica*. Other teleosts are found at similarly high pH (Lahontan cutthroat trout, *Oncorhynchus clarki henshawi*, Wilkie & Wood 1996); salinity (pearl mullet, *Chalcalburnus tarichi;* Sari 2008); low oxygen levels (air breathing gulf toadfish, *Opsanus beta*; Walsh *et al.* 2001; obligatory air-breathers lungfish, *Protopterus* spp., Chapman 2015; catfish *Clarias*, which have a specialised air-breathing organ, Bevan & Kramer 1987); temperature (desert pupfish, *Cyprinodon*, desert gobies, *Chlamydogobius;* Hillyard & Podrabsky 2015). However, no other teleosts appear to experience all the conditions concurrently, and survive large-extent diurnal variations in these variables (Wood *et al.* 2011).

Temperature and osmoregulation

Early experiments in the 1970s explored the tolerance of *A. grahami* to several variables, establishing 24-hr tolerance ranges of 16-40ºC, pH 5-11, salinity maxima 4% and DO from little as 1.1 mg/L (Reite *et al.* 1974). Some acclimatisation was required between these extremes, for example a sudden change from 22ºC to 36ºC proved lethal, even though the fish had originally been collected from lagoon water of 35-38ºC. These experiments support observations from the field, where fish have been observed grazing in waters up to 40ºC but turning back at higher temperatures, with a lethal limit of 42ºC (Trewavas 1983). Although some observations recorded of fish in waters up to 44ºC, it seems that such high temperatures were only experienced transiently in moving between grazing sites (Coe 1966; Albrecht *et al.* 1968). At the lower end of the scale, while *Alcolapia* survived at temperatures <20ºC, initial observations suggested that they would not breed below 30ºC (Coe 1966), although the translocated population in Lake Nakuru experiences significantly lower temperatures of 19-35ºC (Trewavas 1983).

Preliminary (small-scale) investigations of ion osmoregulation in *Alcolapia* suggested that they actively regulate blood ion concentration. *Alcolapia alcalica* exhibited higher levels of Na^+ than ambient water and *A. grahami* samples from Lake Magadi and Nakuru exhibited similar blood concentrations, despite differences in lake ionic content, while conversely the freshwater species *Oreochromis leucostictus* exhibited a different pattern of higher ionic concentration than that of the ambient water (Leatherland *et al.* 1974). Later experiments indicated that *A.*

grahami exhibited low branchial permeability to Na^+ and Cl^-, a feature more common to marine teleosts, although the turnover rates of both ions were considerably lower than that observed in marine fishes (Eddy *et al.* 1981). The blood of *A. grahami* reflects that of the ambient water with a high osmolarity and variable high levels of circulating pH (7.8-8.4), which is thought to be maintained by low pH sensitivity (Bohr effect) of haemoglobin, ensuring the functionality of haemoglobin in the face of fluctuating pH (Lykkeboe *et al.* 1975; Narahara *et al.* 1996). It has further been postulated that oxygen and temperature sensitivity of haemoglobin allow the fish to survive in variable dissolved oxygen levels of negligible oxygen through to super-saturated levels of oxygen (greater than air content), by having a high binding affinity that is decreased at high temperatures, on the premise that high binding affinity would be of benefit in low oxygen situation but in high ambient temperatures (most likely to be during daylight hours when photosynthesis is occurring and DO levels increase) the affinity decreases and oxygen decoupling is more efficient (Lykkeboe *et al.* 1975). Various mechanisms governing acid-base regulation have been suggested, including differential exchange rates of charged compounds (Na^+/H^+, $Na^+/(NH_4)$ and Cl^-/HCO) across the gill epithelium (Johnston *et al.* 1983), and low HCO_3^- permeability and electrogenic HCO_3^- extrusion across the gills (Wood *et al.* 2011).

Urea regulation

The extreme waters of Lake Magadi would prove fatal to most teleost species through neurotoxicity, as they would be unable to excrete ammonia at high alkalinity. Closely related *Oreochromis niloticus* placed in Magadi water die very quickly, but are able to survive transferral to tapwater of the same pH, suggesting an effect of the carbonate alkalinity rather than just pH (Wright *et al.* 1990). An essential adaptation is *A. grahami*'s ability to excrete all nitrogenous waste as urea rather than ammonia (Walsh *et al.* 2001). Most teleost fish are ammoniotelic, such as *O. niloticus* that excretes >80% waste as ammonia, whereas *A. grahami* is ureotelic and excretes all nitrogenous waste as urea (Randall *et al.* 1989). The evolution of complete urotelism is thought to be via the retention of embryonic characteristics, as the enzymes of the ornithine urea cycle and urea transporters are present in teleost embryos to excrete nitrogenous waste before the gills develop sufficiently to excrete ammonia (Wright & Fhyn 2001).

Rates of urea synthesis and excretion in *Alcolapia* are near mammalian values (Narahara *et al.* 1996; Walsh *et al.* 2001). Only 20% of the urea is excreted from the

posterior end of the fish (Wood *et al.* 1994), so it appears that urinary excretion plays only a minor role and that the bulk of urea excretion occurs elsewhere. Gill (branchial) urea permeability has been estimated at 4.74 x 10^{-5} cm^{-1}, which is 10 times higher than a typical lipid bilayer (Walsh *et al.* 2001). Morphological study of the gill ultrastructure suggested rich Golgi apparatus and vesicular trafficking in the gill pavement cells, which may be the manifestation of the facilitated diffusion urea-transport process (Walsh *et al.* 2001).

The level of liver activity of the first enzyme in the urea cycle pathway, carbamoyl-phosphate synthetase III (CPSase III), is too low to account for the observed high rates of urea excretion, and all urea cycle enzymes are present in the muscle tissue of *A. grahami* at levels sufficiently high to account for the high levels of urea excretion (Lindley *et al.* 1999). It has been suggested that the adaptation to ureogenesis in *A. grahami* required a change in organ localisation of the urea cycle enzymes (i.e., muscle as primary site of urea cycle enzyme activity) to allow for the increased rate of urea production (Lindley *et al.* 1999). Large bicarbonate and pH differences exist between *A. grahami* blood plasma levels and that of the surrounding lake waters – it may also be that ureogenesis plays an important role in removing bicarbonate, and so reducing alkalinisation (Randall *et al.* 1989).

Additional adaptations

Additional adaptations to life in highly alkaline waters with fluctuating oxygen levels include: a high intra- and extracellular pH (Wood *et al.* 2002a); a trifurcated oesophagus that permits alkaline water to bypass the acidic stomach when drinking (Bergman *et al.* 2003); facultative air breathing via the air-bladder (Maina 2000), a behaviour that is maintained even in normoxic and hyeroxic conditions (Johannsson *et al.* 2014); unusually large pineal glands, thought to be an adaptation to increases melatonin secretion to protect against extremes of osmolarity, pH and temperature (Relkin 1989); and an extremely high metabolic rate, the highest recorded in poikilothermic teleosts (Narahara *et al.* 1996; Wood *et al.* 2002a). Around half of the increased metabolic rate is thought to be due to the requirements for acid-base regulation in high pH water, with improved swimming performance seen in diluted water (Wood *et al.* 2002a). Air-breathing allows survival in extremely low levels of DO, with studies of Lake Nakuru *A. grahami* suggesting that levels of DO lower than 1.36 mg O2/L induce air-breathing behaviour, and sufficient oxygen was attained by this method for the fish to survive >2 hours (and survive when replaced in aerated water) (Schwan & Lamberti 1986). Despite the specialised adaptations to the

extreme environment, changes in osmolarity are tolerated, with survival of acute changes to water of 10% Magadi lakewater concentration (based on Cl^- concentration), and surviving gradual acclimation to 1% and 200-250% (Wood *et al.* 2002b; 2011).

Alcolapia grahami was introduced to Lake Nakuru, Kenya, in the 1950s (Hickley *et al*. 2008) in an attempt to control the mosquito population. Lake Nakuru is also a soda lake (pH ~10) but with a lower water temperature than Lake Magadi (Vareschi 1979). Despite little evidence of success in the intended biological control of mosquito larvae, the introduction provides an interesting natural experiment in the adaptation of non-native fish species to new environments. Specimens from this lake are now reported to grow to twice the size of individuals in Lake Magadi (Vareschi 1979), and exhibit a noted difference in body shape (pers. obs.), so it appears that the extreme conditions of the native Magadi waters may constrain the growth of *A. grahami*.

Alcolapia biology and reproduction

All *Alcolapia* species exhibit sexual dichromatism, nest construction, external fertilisation, and maternal mouthbrooding (Trewavas 1983; Seegers & Tichy 1999). There is no distinct breeding period, with spawning occurring year-round; observations of *A. grahami* record the brooding period as ~2 weeks with males reported to mature and mate within 7 weeks of release from the maternal mouth (Trewavas 1983; Coe 1966). The fish undergo continual spermatogenesis, presumably to maintain continuous breeding, and the spermatozoon appears to show adaptations to the extreme environment (Papah *et al.* 2013).

Aquarium populations of *A. grahami* have been sustained by some researchers, although descriptions detail only observations of behaviour rather than behavioural or mate-choice experiments (Albrecht 1968; Seegers & Tichy 1999). Laboratory populations have been maintained up to F6 (Seegers *et al.* 2001) and successfully bred in freshwater (Albrecht *et al.* 1968). No recent reports of laboratory-kept *Alcolapia* have appeared in the literature, but anecdotal accounts of *A. alcalica* breeding populations are reported from the aquarium trade (in particular, online cichlid breeding fora), where they are reported to thrive and breed even in freshwater or temperate tank conditions (i.e., non-extreme conditions), where a life span of 7-8 years is recorded.

Alcolapia are observed to be constant grazers, with almost continual activity in daylight hours (Coe 1966; Johannsson *et al.* 2014). Primary diet, based on stomach

contents analysis, is mainly cyanobacteria (90%) with a small proportion of crustacea and larvae (Coe 1969). The characteristic trophic morphology in each of the species is maintained in aquaria populations fed on the same food (Seegers *et al.* 2001). All species cannabilistically target eggs and fry and will snatch eggs from brooding females if possible (Woodhouse 1912; Coe 1966; Albrecht 1968). When kept in a mixed species tank, *A. alcalica* repeatedly targeted breeding *O. niloticus* to steal eggs on release from the female despite being considerably smaller than the breeding pair (Seegers *et al.* 2001). Females have been observed to swim to shallow pools to release mature fry, thought to be an adaptation to release the young fish in less extreme conditions, but also prevent conspecific predation, as even the mother will target the fry as food within a few hours of release (Coe 1969). In other cichlid species, for example *Pseudocrenilabrus*, the mouthbrooding period has shown to be substantially shorter (30%) in conditions of low oxygen, suggesting a metabolic cost to mouthbrooding (Reardon & Chapman 2010). The large extent of trophic variability in *Alcolapia* is also thought to impact on mouth brooding capability, as the species with sub-terminal or inferior mouths and shorter lower jaws (*A. latilabris* and *A. ndalalani*) brood fewer offspring at one time (~50%) than the terminal mouth *A. alcalica* (Coe 1969; Tichy & Seegers 1999)

The phylogenetic relationship of *Alcolapia* and *Oreochromis*

Tribe Oreochromini

Alcolapia nests within the 'tilapia' group, which includes *Tilapia, Sarotherodon* and *Oreochromis* (Trewavas 1983). Additional distinctions of these groups have been made based on parental care, as *Tilapia* are substrate brooders while *Sarotherodon* and *Oreochromis* are both mouthbrooders (Trewavas 1983), although the derived mouthbrooding behaviour has been shown to have several independent origins (Klett & Meyer 2002; Mank *et al.* 2005). Recent molecular work has divided the tilapiine group further, with the genus *Tilapia* placed in the Tilapiini tribe, and the genera *Sarotherodon* and *Oreochromis* together placed in the Oreochromini (Schwarzer *et al.* 2009; Dunz & Schliewen 2013).

The Oreochromini and *Oreochromis* represent recent cichlid radiations, with divergence estimates of 12.8-21.4 MYA and 6.4-9.7 MYA for the tribe and genus respectively (Won *et al.* 2006; Schwarzer *et al.* 2009). These estimates were based on multigene datasets combining nuclear and mtDNA loci, and it is notable that previous estimates for *Oreochromis* divergence based on single mtDNA markers delivered significantly more recent divergence times of (~3.1 MYA; Nagl *et al.* 2001).

However, mtDNA may be incongruent with nuclear data and is most likely insufficient for dating analysis (Ballard & Whitlock 2004; Won *et al.* 2006). The earliest *Oreochromis* fossils (all extinct species) are from the Late Miocene / Early Pliocene and dates are consistent with the multigene dataset estimates (and used by Schwarzer *et al*. (2009) as a fossil calibration): *O. harissae* 4.3-4.4 Ma (Murray & Stewart 1999); *O. lorenzoi* ~6.0 Ma (Carnevale *et al.* 2003); *O. martyni* 9.3-12 Ma (Van Couvering 1982; Murray 2001).

Oreochromis diversity

While not nearly as species rich as the freshwater lacustrine cichlid genera such as *Haplochromis* (Lowe-McConnell 1959; Trewavas 1982a; Klett & Meyer 2002; Seehausen 2007), *Oreochromis* are widespread and are not restricted by water chemistry. Some authors have suggested that the lower diversity of the tilapiine cichlids is due to a generalised condition and subsequent ability to adapt without diversification (Fryer & Iles 1969), although other researchers refute this argument and hypothesise that mouthbrooding and trophic adaptations indicate that the group were already highly specialised, so would not have had natural competition in new environments and thus diversification would not have been stimulated (Trewavas 1982b; 1983). The tilapias have typically been considered to exhibit allopatric speciation based on their more widespread distribution in rivers, but cases of sympatric speciation in *Oreochromis* have been reported in both lakes and rivers (Trewavas 1983; Seehausen 2007).

Adaptation to soda conditions

Of the 33 described *Oreochromis* species, 11 can tolerate or occur in soda-like conditions (Table 1.4). The typically freshwater *O. niloticus* and *O. mossambicus* have successfully colonised salt lakes (Riedel & Costa-Pierce 2005) and thermal hot springs (Bezault *et al.* 2007), while *O. amphimelas* occurs in the springs and main water bodies of soda lakes in Tanzania. *Oreochromis amphimelas* is the only cichlid found in soda Lakes Eyasi and Manyara and co-occurs with *O. niloticus* and *O. esculentus* in Lakes Singida and Kitangiri. It has been suggested that *O. amphimelas* might be the closest relative to *Alcolapia* based on adaptation to soda conditions (although *O. amphimelas* experiences less extreme conditions) (Trewavas 1983). There is also a hydrological connection between the two basins, with water flowing northwards from Manyara to Natron and the lowest border of Manyara forming an overspill to Natron (Hillaire-Marcel & Casanova 1987; Bachofer *et al.* 2014), however the suggestion that the two basins were joined in a palaeolake as recently as 10 KYA (Holdship 1976) has not been supported by geological

evidence (Casanova & Hillaire-Marcel 1992). Like Magadi, Lake Manyara also seems to have been inhabited by a larger freshwater cichlid in a previous palaeolake based on fossils (Schluter *et al.* 1992). However, unlike Natron-Magadi, Manyara contains only one species (*O. amphimelas*).

Morphology also supports a close relationship of *Alcolapia* to *O. amphimelas*, although there are also notable differences, which prevented them being placed in the same sub-genus by Trewavas (1983). While molecular analyses have consistently recovered the monophyly of *Alcolapia*, the relationship to *Oreochromis* is unresolved, with *O. amphimelas* (Seegers *et al.* 1999; Nagl *et al.* 2001), *O. malagarasi* (Seegers *et al.* 1999), *O. tanganicae* (Schwarzer *et al.* 2009) and *O. variabilis* (Kavembe *et al.* 2013) being proposed as possible sister species to the *Alcolapia* species flock. Other than *O. amphimelas,* the other three species are all found in freshwater. Phylogenetic relationships within *Alcolapia* are also unresolved, with shared mtDNA haplotypes between all species and from both lakes (Natron and Magadi) (Seegers *et al*. 1999). Previous molecular studies of *Alcolapia* are reviewed in chapter three.

Sex determination

The karyotype evolution of the East African cichlids is generally conservative. *Oreochromis* species all have 44 chromosomes, with the exception of *O. karongae*, which has 38 (Harvey *et al.* 2002; Mota-Velasco *et al.* 2010), whereas *Alcolapia* have 48 chromosomes (n=24) (Poletto *et al.* 2010). The mechanism of sex determination in *Alcolapia* is not known, although the mechanisms in *Oreochromis* species include female heterogametic WZ-ZZ system (*O. karongae*), male heterogametic XX-XY system (*O. niloticus*), and autosomal sex-determining loci (*O. aureus* and *O. mossambicus*) (Mair *et al.* 1991; Cnaani *et al.* 2008). Environmental conditions also influence sex determination, with temperature having a major impact (Palaiokostas *et al.* 2013), and naturally sex reversed individuals have been reported in wild populations (Bezault *et al.* 2007; Baroiller *et al.* 2009).

Trewavas (1983) notes sexual dimorphism (as well as sexual dichromatism) in *Oreochromis* and *Alcolapia* species, however, this only relates to soft-tissue or internal features: genital papilla; male elongation of soft dorsal and anal fins, and simplification of oral teeth in males. In some *Oreochromis* species, there is an elongation of the jaws in male specimens (Trewavas 1983).

Table 1.4 *Alcolapia* and *Oreochromis* adaptation to soda conditions.

Temp: temperature range; Soda: maximum salinity and/or pH in which species occur.
Citations for original species descriptions are included in the reference list. Subscript letters following the species name distinguish multiple publications of a single author in the same year.

	Species	Distribution	Temp	Soda
Alcolapia	*A. alcalica* (Hilgendorf 1905)	Natron, TZA	20-42	>40ppt, pH 11
	A. grahami (Boulenger 1912)[a]	Magadi, Nakuru, KEN	20-42	>40ppt, pH 11
	A. latilabris (Seegers & Tichy 1999)	Natron, TZA	20-42	>40ppt, pH 11
	A. ndalalani (Seegers & Tichey 1999)	Natron, TZA	20-42	>40ppt, pH 11
Oreochromis	*O. amphimelas* (Hilgendorf 1905)	Soda lakes, TZA	20-30	58 ppt
	O. andersonii (Castelnau 1861)	South-central Africa	18-33	20ppt
	O. angolensis (Trewavas 1973)	Southern Africa	-	-
	O. aureus (Steindachner 1864)	Eurasia, Africa, USA[†]	12-32	45 ppt,
	O. chungruruensis (Ahl 1924)	Chungruru, TZA	-	-
	O. esculentus (Graham 1928)	Nile, East African Lakes	23-29	-
	O. hunteri (Günther 1889)	Chala, TZA	-	-
	O. ismailiaensis (Mekkaway 1995)	EGY	-	-
	O. jipe (Lowe 1955)	Jipe, Pangani, TZA	-	-
	O. karomo (Poll 1948)	Tanganyika, E. Africa	-	-
	O. karongae (Trewavas 1941)	Malawi, E. Africa	-	-
	O. korogwe (Lowe 1955)	Eastern Africa	-	-
	O. lepidurus (Boulenger 1899)	Central Africa	-	-
	O. leucostictus (Trewavas 1933)	Edward, George, UGA	15-38	pH 7-9
	O. lidole (Trewavas 1941)	Malawi, Chungruru, TZA	-	-
	O. macrochir (Boulenger 1912)[b]	S. Africa, Hawaii (introduced)	18-32	20ppt
	O. malagarasi (Trewavas 1983)	Eastern Africa	-	-
	O. mortimeri (Trewavas 1966)[b]	Southern Africa	19-32	-
	O. mossambicus (Peters 1852)	SE Africa, widely introduced	17-35	100 ppt
	O. mweruensis (Trewavas 1983)	Congo river system	-	-
	O. niloticus baringoensis (Trewavas 1983)	Baringo, KEN	-	-
	O. niloticus cancellatus (Nichols 1923)	Awash basin, ETH	17-26	-
	O. niloticus eduardianus (Boulenger 1912)[b]	Edward, UGA	-	-
	O. niloticus filoa (Trewavas 1983)	Hot springs, Awash, ETH	32-39	
	O. niloticus niloticus (Linnaeus 1758)	NE Africa, widely introduced	14-32	30 ppt
	O. niloticus sugutae (Daget 1991)	Karpeddo soda springs,	20-38	-
	O. niloticus tana (Seyoum 1992)	Lake Tana, ETH	-	-
	O. niloticus vulcani (Trewavas 1933)	Crater lake, Turkana, KEN	-	-
	O. placidus placidus (Trewavas 1941)	Southeastern Africa	-	'freshwater'
	O. placidus ruvumae (Trewavas 1966)	Upper Ruvuma, SE Africa	-	-
	O. rukwaensis (Hilgendorf 1903)	Lake Rukwa, TZA	-	-
	O. saka (Lowe 1953)	Lake Malawi, East Africa	-	-
	O. salinicola (Poll 1948)	Central Africa	-	25-35 ppt
	O. schwebischi (Sauvage 1884)	West-Central Africa	-	-
	O. shiranus chilwae (Trewavas1966)[a]	Lake Chilwa, MWI	21-37	30 ppt
	O. shiranus shiranus (Boulenger 1896)	Lake Malawi and drainage	-	-
	O. spilurus niger (Günther 1894)	Kibwezi River, KEN	19-32	-
	O. spilurus percivali (Boulenger 1912)	Hot springs, KEN	20-38	'alkaline'
	O. spilurus spilurus (Günther 1894)	KEN, introduced elsewhere	20-31	-
	O. squamipinnis (Günther 1864)	Lake Malawi	-	-
	O. tanganicae (Günther 1893)	Lake Tanganyika	-	-
	O. upembae (Thys van den Audenaerde 1964)	Congo river basin	-	-
	O. urolepis hornorum (Trewavas	TNZ	-	-
	O. urolepis urolepis (Norman 1922)	TNZ	25-38	pH 8.4
	O. variabilis (Boulenger 1906)	Lake Victoria and drainage	23-28	-

Note that the temperatures and salinity/soda conditions in Table 1.4 are those at which the species have been recorded to naturally occur. Several studies have shown many of these species are able to tolerate/survive higher levels in laboratory conditions (though fewer have explored successful reproduction at extreme conditions; e.g., Morgan 1972). All ranges and maxima for temperature and soda conditions are based on Trewavas (1982a; 1983); Bezault *et al.* (2007).

Study hypotheses and research questions

Soda lake cichlids represent an excellent system for investigating processes generating biological diversity, since the system is young geologically (Eugster 1986), spatially restricted, and diversity is limited compared to the African Great Lakes. This thesis seeks to characterise the evolutionary relationships of *Alcolapia* and address the question of whether *Alcolapia* represent an adaptive radiation.

Chapter two introduces the methods used in this thesis and their application to cichlid species research. Chapter three characterises *Alcolapia* evolutionary relationships through dense sampling employing a genome-wide SNP approach. As well as phylogenomically testing species hypotheses, population connectivity of *Alcolapia* within each of the lakes and interspecific gene flow within the radiation is investigated. Chapter four examines the distribution of differentiation heterogeneity across the genome, and tests for loci under selection between species, populations, and sexes. Chapter five uses stable isotope analysis (as a proxy for niche space) and geometric morphometrics of body shape and pharyngeal jaw to test the prediction that ecological speciation has been an important driver of adaptive diversification in the soda lake cichlids. Ecomorphology of the species flock is considered within a phylogenetic framework to test the degree of niche partitioning between sympatrically occurring species, whether morphological adaptation is correlated with ecological niche utilisation; and the correlation of ecological and morphological differentiation with phylogeny. Chapter six considers variation of intraspecific ecomorphology, tests whether the ubiquitous *A. alcalica* shows a greater extent of variability than the species constrained to southern Natron and whether local adaptation (and adaptive divergence) is observed in the face of high levels of gene flow. Comparisons of allopatric (monospecific) and sympatric populations are made to consider competition effects on niche exploitation and whether ecological niche is conserved in translocated populations. Finally, chapter seven draws conclusions and outlines future research goals.

References

Albrecht H (1968) Freiwasserbeobachtungen an Tilapien (Pisces, Cichlidae) in OstAfrika. *Zeitschrift für Tierpsychologie*, **25**, 377–394.

Albrecht H, Apfelbach R, Wickler W (1968) Über die Eigenständigkeit der Art *Tilapia grahami* Boulenger, ihren Grubenbau und die Zucht in reines Susswasser (Pisces, Cichlidae). *Senckenberg Biologica*, **49**, 107–118.

Ahl E (1924) Über einen neuen Cichliden aus Ost-Afrika. *Zoologischer Anzeiger - A Journal of Comparative Zoology*, **59**, 86–87.

Bachofer F, Quénéhervé G, Märker M (2014) The delineation of paleo-shorelines in the Lake Manyara basin using TerraSAR-X data. *Remote Sensing*, **6**, 2195–2212.

Baldwin BG, Sanderson MJ (1998) Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proceedings of the National Academy of Sciences*, **95**, 9402–9406.

Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.

Baroiller J-F, D'Cotta H, Saillant E (2009) Environmental effects on fish sex determination and differentiation. *Sexual Development*, **3**, 118–135.

Barrett R, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, **23**, 38–44.

Behr H-J (2002) Magadiite and Magadi chert: A critical analysis of the silica sediments in the lake Magadi basin, Kenya. In: *Sedimentation in Continental Rifts*, pp. 257–273. Soceity for Sedimentary Geology.

Bergman AN, Laurent P, Otiang'a-Owiti G *et al.* (2003) Physiological adaptations of the gut in the Lake Magadi tilapia, *Alcolapia grahami*, an alkaline- and saline-adapted teleost fish. *Comparative Biochemistry and Physiology, Part A*, **136**, 701–715.

Bevan DJ, Kramer DL (1987) The respiratory behaviour of an air-breathing-catfish, *Clarias macrocephalus* (Clariidae). *Canadian Journal of Zoology*, **65**, 348–353.

Bezault E, Feder C, Derivaz M *et al*. (2007) Sex determination and temperature-induced sex differentiation in three natural populations of Nile tilapia (*Oreochromis niloticus*) adapted to extreme temperature conditions. *Aquaculture*, **272**, S3–S16.

Bezault E, Mwaiko S, Seehausen O (2011) Population genomic tests of models of adaptive radiation in Lake Victoria region cichlid fish. *Evolution*, **65**, 3381–3397.

Birand A, Vose A, Gavrilets S (2012) Patterns of species ranges, speciation, and extinction. *The American Naturalist*, **179**, 1–21.

Boulenger GA (1896) Descriptions of new fishes from the Upper Shiré River, British Central Africa, collected by Dr. Percy Rendall, and presented to the British Museum by Sir Harry H. Johnston, K.C.B. *Proceedings of the Zoological Society of London*, **1896**, 915–920.

Boulenger GA (1899) A revision of the African and Syrian fishes of the family Cichlidae. *Proceedings of the Zoological Society of London*, **2**, 98–143.

Boulenger GA (1906) Descriptions of new fishes discovered by Mr. E. Degen in Lake Victoria. *Annals and Magazine of Natural History*, **17**, 433–452.

Boulenger GA (1912a) Description of a new fish from British East Africa. *Annals and Magazine of Natural History*, **9**, 519–521.

Boulenger GA (1912b) Descriptions of three new African cichlid fishes of the genus *Tilapia*, preserved in the British Museum. *Annals and Magazine of Natural History*, **10**, 138–140.

Brawand D, Wagner CE, Li YI *et al.* (2014) The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, **513**, 375–381.

Burrough SL, Thomas DSG (2009) Geomorphological contributions to palaeolimnology on the African continent. *Geomorphology*, **103**, 285–298.

Butzer KW, Isaac GL, Richardson JL, Washbourn-Kamau C (1972) Radiocarbon dating of East African Lake Levels. *Science*, **175**, 1069–1076.

Carnevale G, Sorbini C, Landini W (2003) *Oreochromis lorenzoi*, a new species of tilapiine cichlid from the Late Miocene of Central Italy. *Journal of Vertebrate Paleontology*, **23**, 508–516.

Casanova J, Hillaire-Marcel C (1992) Chronology and paleohydrology of Late Quaternary high lake levels in the Manyara basin (Tanzania) from isotopic data (^{18}O, ^{13}C, ^{14}C, Th/U) on fossil stromatolites. *Quaternary Research*, **38**, 205–226.

Castelnau FL (1861) *Mémoire sur les poissons de l'Afrique australe*. Paris.

Chapman LJ (2015) Low-oxygen lifestyles. In: *Extremophile Fishes* (Eds. Riesch R, Tobler M, Plath M), pp. 9–33. Springer International, Switzerland.

Cnaani A, Lee BY, Zilberman N *et al.* (2008) Genetics of sex determination in tilapiine species. *Sexual Development*, **2**, 43–54.

Coe MJ (1966) The Biology of *Tilapia grahami* Boulenger in Lake Magadi, Kenya. *Acta Tropica*, **23**, 146–198.

Coe MJ (1969) Observations on *Tilapia alcalica* Hilgendorf an endemic cichlid fish from Lake Natron, Tanzania. *Revue de Zoologie et de Botanique Africaines*, **80**, 1–9.

Copley H (1958) *Common freshwater fishes of East Africa*. Witherby, London.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, MA.

Daget J, Grosse J-P, Teugels GG, Thys van den Audenaerde DFE (1991) *Check-list of the freshwater fishes of Africa*. CLOFFA. MRAC, Tervuren, Belgium.

Darwin C (1859) *On the origin of species by means of natural selection, or the preservationof favoured races in the struggle for life*. John Murray, London.

Dieckmann U, Doebeli M (1999) On the origin of species by sympatric speciation. *Nature*, **400**, 354–357.

Dunz AR, Schliewen UK (2013) Molecular phylogeny and revised classification of the haplotilapiine cichlid fishes formerly referred to as "Tilapia." *Molecular Phylogenetics and Evolution*, **68**, 64–80.

Eddy FB, Bamford OS, Maloiy GMO (1981) Na^+ and Cl^- effluxes and ionic regulation in *Tilapia grahami*, a fish living in condtions of extreme alkalinity. *Journal of Experimental Biology*, **91**, 349–353.

Elmer KR, Fan S, Gunter HM *et al.* (2010a) Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes. *Molecular Ecology*, **19**, 197–211.

Elmer KR, Fan S, Kusche H *et al.* (2014) Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel. *Nature Communications*, **5**, 1–8.

Elmer KR, Lehtonen TK, Kautt AF, Harrod C, Meyer A (2010b) Rapid sympatric ecological differentiation of crater lake cichlid fishes within historic times. *BMC Biology*, **8**, 60.

Eschmeyer WN (Ed.) (2015) *Catalog of fishes electronic version. Updated 7 January 2015*. California Academy of Sciences.

Eugster HP (1986) Lake Magadi, Kenya: a model for rift valley hydrochemistry and sedimentation? *Geological Society, London, Special Publications*, **25**, 177–189.

Fan S, Elmer KR, Meyer A (2011) Genomics of adaptation and speciation in cichlid fishes: recent advances and analyses in African and Neotropical lineages. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 385–394.

Fischer JG (1883) Über einige afrikanische Reptilien, Amphibien und Fische des Naturhistorischen Museums. *Jahrbuch der Hamburgischen Wissenschaftlichen Anstalten*, **1**, 1–39.

Fitzpatrick BM, Fordyce JA, Gavrilets S (2008) What, if anything, is sympatric speciation? *Journal of Evolutionary Biology*, **21**, 1452–1459.

Ford AGP, Dasmahapatra KK, Rüber L *et al.* (2015) High levels of interspecific gene flow in an endemic cichlid fish adaptive radiation from an extreme lake

environment. *Molecular Ecology*, doi: 10.1111–mec.13247.

Fryer G, Iles TD (1969) Alternative routes to evolutionary success as exhibited by African cichlid fishes of the genus *Tilapia* and the species flocks of the Great Lakes. *Evolution*, **23**, 359–369.

Fryer G, Iles TD (1972) *The cichlid fishes of the Great Lakes of Africa: their biology and evolution*. Oliver & Boyd, Edinburgh.

Gavrilets S, Losos JB (2009) Adaptive radiation: Contrasting theory with data. *Science*, **323**, 732–737.

Geiger MF, McCrary JK, Stauffer JR Jr. (2010) Description of two new species of the Midas cichlid complex (Teleostei: Cichlidae) from Lake Apoyo, Nicaragua. *Proceedings of the Biological Society of Washington*, **123**, 159–173.

Givnish TJ (1997) Adaptive radiation and molecular systematics: issues and approaches. In: *Molecular evolution and adaptive radiation* (eds Givnish TJ, Sytsma KJ), pp. 1–54. Cambridge University Press, Cambridge.

Givnish TJ (2015) Adaptive radiation versus 'radiation' and 'explosive diversification': why conceptual distinctions are fundamental to understanding evolution. *New Phytologist*, doi: 10.1111–nph.13482.

Graham M (1928). *Tilapia esculenta*, a new species of cichlid fish from the Victoria Nyanza, with notes on the habits of this species and the nearly-related *Tilapia variabilis*, Blgr. *Journal of Natural History Series 10*, **2**, 209–213.

Grant PR, Grant BR (2008) *How and why species multiply: the radiation of Darwin's finches*. Princeton Univ. Press, Princeton, NJ.

Grant WD (2006) Alkaline environments and biodiversity. In: *Extremophiles (Life under Extreme Environmental Conditions)* (Eds. Gerday C, Glansdorff N), pp. 1–20. UNESCO, Oxford.

Greenwood PH (1974) Cichlid fishes of Lake Victoria, East Africa: Biology and evolution of a species flock. *Bulletin of the British Museum (Natural History) Zoology*, **6**, 1–134.

Günther A (1864) Report on a collection of reptiles and fish made by Dr Kirk in the Zambesi and Nyssa regions. *Proceedings of the Zoological Society of London*, **1864**, 303–314.

Günther AE (1889) On some fishes from the Kilima-njaro district. *Proceedings of the Zoological Society of London*, **1889**, 70–72.

Günther AE (1894) Report on the collection of reptiles and fishes made by Dr. J. W. Gregory, during his expedition to Mount Kenia. *Proceedings of the Zoological Society of London*, **1894**, 84–91.

Habel JC, Husemann M, Schmitt T *et al.* (2012) Microallopatry caused strong diversification in *Buthus* scorpions (Scorpiones: Buthidae) in the Atlas Mountains (NW Africa). *PLoS ONE*, **7**, e29403.

Harmon LJ, Schulte JA, Larson A, Losos JB (2003) Tempo and mode of evolutionary radiation in Iguanian lizards. *Science*, **301**, 961–965.

Harvey SC, Campos-Ramos R, Kennedy DD *et al.* (2002) Karyotype evolution in Tilapia: mitotic and meiotic chromosome anlaysis of *Oreochromis karongae* and *O. niloticus* x *O. karongae* hybrids. *Genetica*, **115**, 169–177.

Hickley P, Muchiri M, Britton R, Boar R (2008) Economic gain versus ecological damage from the introduction of non-native freshwater fish: Case studies from Kenya. *The Open Fish Science Journal*, **1**, 36–46.

Hilgendorf F (1905) Fische von Deutsche und Englisch Ost-Afrika. Gesammelt von Oscar Neumann 1893-1895. *Zoologische Jahrbücher*, **22**, 405–420.

Hilgendorf FM, Pappenheim P (1903) Über die Fischfauna des Rukwa-Sees. *Sitzungsberichte der Gesellschaft Naturforschender Freunde zu Berlin*, **6**, 259–271.

Hillaire-Marcel C, Casanova J (1987) Isotopic hydrology and paleohydrology of the Magadi (Kenya) - Natron (Tanzania) basin during the late Quaternary. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **58**, 155–181.

Hillyard SD, Podrabsky JE (2015) Desert environments. In: *Extremophile Fishes* (Eds. Riesch R, Tobler M, Plath M), pp. 59–83. Springer International, Switzerland.

Holdship S (1976) *The paleolimnology of Lake Manyara, Tanzania: A diatom analysis of a 56 meter sediment core*. Duke University.

Hoy RD, Stephens RD (1979) *Field study of lake evaporation: analysis of data from phase 2 storages and summary of phase 1 and phase 2*. Australian Government Publishing Service, Canberra.

Hudson AG, Vonlanthen P, Seehausen O (2011) Rapid parallel adaptive radiations from a single hybridogenic ancestral population. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 58–66.

Hulsey CD, Hollingsworth PR Jr, Fordyce JA (2010) Temporal diversification of Central American cichlids. *BMC Evolutionary Biology*, **10**, 279.

Johannsson OE, Bergman HL, Wood CM *et al.* (2014) Air breathing in Magadi tilapia *Alcolapia grahami*, under normoxic and hyperoxic conditions, and the association with sunlight and reactive oxygen species. *Journal of Fish Biology*, **84**, 844–863.

Johnston IA, Eggy FB, Maloiy GMO (1983) The effects of temperature on muscle pH, adenylate and phosphogen concentrations in *Oreochromis alcalicus grahami*, a fish adapted to the alkline hot-spring. *Journal of Fish Biology*, **23**, 717–724.

Jones BE, Grant WD, Duckworth AW, Owenson GG (1998) Microbial diversity of soda lakes. *Extremophiles*, **2**, 191–200.

Jones BF, Eugster HP, Rettig SL (1977) Hydrochemistry of the Lake Magadi basin, Kenya. *Geochimica et Cosmochimica Acta*, **41**, 53–72.

Kadigi RMJ, Kilima FT, Kashaigili JJ (2012) *A comparative study of costs and benefits of soda ash mining and promotion of ecotourism and sustainable use of natural resources in Lake Natron basin, Tanzania*. Birdlife International, Birdlife Africa Partnership Secretariat, Nairobi, Kenya.

Kahilainen KK, Patterson WP, Sonninen E *et al*. (2014) Adaptive radiation along a thermal gradient: Preliminary results of habitat use and respiration rate divergence among whitefish morphs. *PLoS ONE*, **9**, e112085.

Kaufman A, Margaritz M, Paul M *et al.* (1990) The ^{36}Cl ages of the brines in the Magadi-Natron basin, East Africa. *Geochimica et Cosmochimica Acta*, **54**, 2827–2833.

Kautt AF, Elmer KR, Meyer A (2012) Genomic signatures of divergent selection and speciation patterns in a “natural experiment,” the young parallel radiations of Nicaraguan crater lake cichlid fishes. *Molecular Ecology*, **21**, 4770–4786.

Kavembe GD, Machado-Schiaffino G, Meyer A (2013) Pronounced genetic differentiation of small, isolated and fragmented tilapia populations inhabiting the Magadi Soda Lake in Kenya. *Hydrobiologia*, **739**, 55–71.

Klett V, Meyer A (2002) What, if anything, is a Tilapia? - Mitochondrial ND2 phylogeny of tilapiines and the evolution of parental care systems in the African cichlid fishes. *Molecular Biology and Evolution*, **19**, 865–883.

Koblmüller S, Schliewen UK, Duftner N *et al.* (2008) Age and spread of the haplochromine cichlid fishes in Africa. *Molecular Phylogenetics and Evolution*, **49**, 153–169.

Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews Genetics*, **5**, 288–298.

Kocher TD, Conroy JA, McKaye KR, Stauffer JR (1993) Similar morphologies of cichlid fish in Lakes Tanganyika and Malawi are due to convergence. *Molecular Phylogenetics and Evolution*, **2**, 158–165.

Kondrashov AS, Kondrashov FA (1999) Interactions among quantitative traits in the course of sympatric speciation. *Nature*, **400**, 351–354.

Kozak KM, Wahlberg N, Neild AFE *et al.* (2015) Multilocus species trees show the recent adaptive radiation of the mimetic *Heliconius* butterflies. *Systematic Biology*, **64**, 505–524.

Lamichhaney S, Berglund J, Almén MS *et al.* (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, **518**, 371–375.

Lande R, Seehausen O, van Alphen JJM (2001) Mechanisms of rapid sympatric speciation by sex reversal and sexual selection in cichlid fish. *Genetica*, **112-113**, 435–443.

Leatherland JF, Hyder M, Ensor DM (1974) Regulation of plasma Na and K^+ concentrations in five African species of *Tilapia* fishes. *Comparative Biochemistry And Physiology A*, **48A**, 699–710.

Lerner HRL, Meyer M, James HF *et al*. (2011) Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of hawaiian honeycreepers. *Current biology,* **21**, 1838–1844.

Liem KF (1973) Evolutionary strategies and morphological innovations: Cichlid pharyngeal jaws. *Systematic zoology*, **22**, 425.

Lindley TE, Scheiderer CL, Walsh PJ *et al.* (1999) Muscle as the primary site of urea cycle enzyme activity in an alkaline lake-adapted tilapia, *Oreochromis alcalicus grahami*. *The Journal of Biological Chemistry*, **274**, 29858–29861.

Linnaeus C (1758) *Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis.* 10th edition. Laurentius Salvius, Stockholm.

Loh YHE, Bezault E, Muenzel FM *et al.* (2013) Origins of shared genetic variation in African cichlids. *Molecular Biology and Evolution*, **30**, 906–917.

Losos JB, Thorpe RS (2004) Evolutionary diversification of Caribbean *Anolis* lizards. In: *Adaptive Speciation* (Eds. Dieckmann U, Doebeli M, Metz JAJ, Tautz D), pp. 322–344. Cambridge University Press, Cambridge.

Lowe RH (1953) Notes on the ecology and evolution of Nyasa fishes of the genus *Tilapia*, with a description of *T. saka* Lowe. *Proceedings of the Zoological Society of London*, **122**, 1035–1041.

Lowe RH (1955) New species of *Tilapia* (Pisces, Cichlidae) from Lake Jipe and the Pangani River, East Africa. *Bulletin of the British Museum (Natural History) Zoology*, **2**, 349–368.

Lowe-McConnell RH (1959) Breeding behaviour patterns and ecological differences between Tilapia species and their significance for evolution within the genus Tilapia (Pisces: Cichlidae). *Proceedings of the Zoological Society of London*, **132**, 1–30.

Lykkeboe G, Maloiy GMO, Johansen K (1975) Functional properties of hemoglobins in the teleost *Tilapia grahami*. *Journal of Comparative Physiology*, **104**, 1–11.

Machado-Schiaffino G, Kautt AF, Kusche H, Meyer A (2015) Parallel evolution in Ugandan crater lakes: repeated evolution of limnetic boy shapes in haplochromine cichlid fish. *BMC Evolutionary Biology*, **15**, 9.

Maina JN (2000) Functional morphology of the gas-gland cells of the air-bladder of *Oreochromis alcalicus grahami* (Teleostei: Cichlidae): an ultrastructural study on a fish adapted to a severe, highly alkaline environment. *Tissue and Cell*, **32**, 117–132.

Mair GC, Scott AG, Penman DJ *et al*. (1991) Sex determination in the genus *Oreochromis*. *Theoretical and Applied Genetics*, **82**, 144–152.

Mank JE, Promislow DEL, Avise JC (2005) Phylogenetic perspectives in the evolution of parental care in ray-finned fishes. *Evolution*, **59**, 1570–1578.

Martin CH, Feinstein LC (2014) Novel trophic niches drive variable progress towards ecological speciation within an adaptive radiation of pupfishes. *Molecular Ecology*, **23**, 1846–1862.

Martin CH, Cutler JS, Friel JP *et al.* (2015) Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. *Evolution*, doi: 10.1111–evo.12674.

Matschiner M, Hanel R, Salzburger W (2011) On the origin and trigger of the notothenioid adaptive radiation. *PLoS ONE*, **6**, e18911.

Maxon RM, Ofcansky TP (2014) *Historical dictionary of Kenya*. Rowman & Littlefield, Lanham.

McKetta JJ (1995) *Encyclopedia of chemical processing and design*. Marcel Dekker, Inc, New York.

McKinnon JS, Mori S, Blackman BK *et al.* (2004) Evidence for ecology's role in speciation. *Nature*, **429**, 294–298.

Mekkawy IAA (1995) Description of *Oreochromis ismailiaensis* sp. n., and its hybrid with *Oreochromis niloticus* (Linnaeus, 1758) (Perciformes; Cichlidae) from Egypt. *Bulletin of the Faculty of Science, Assiut University, Egypt*, **24**, 1–27.

Meyer BS, Indermaur A, Ehrensperger X *et al.* (2015) Back to Tanganyika: a case of recent trans-species-flock dispersal in East African haplochromine cichlid fishes. *Royal Society Open Science*, **2**, 140498–140498.

Mikhodyuk OS, Gerasimenko LM, Akimov VN *et al*. (2008a) Ecophysiology and polymorphism of the unicellular extremely natronophilic cyanobacterium Euhalothece sp. Z-M001 from Lake Magadi. *Microbiology*, **77**, 717–725.

Mikhodyuk OS, Zavarzin GA, Ivanovsky RN (2008b) Transport systems for carbonate in the extremely natronophilic cyanobacterium *Euhalothece* sp. *Microbiology*, **77**, 412–418.

Morgan PR (1972) Causes of mortality in the endemic *Tilapia* of Lake Chilwa (Malawi). *Hydrobiologia*, **40**, 101–119.

Mota-Velasco JC, Ferreira IA, Cioffi MB *et al.* (2010) Characterisation of the chromosome fusions in *Oreochromis karongae*. *Chromosome Research*, **18**, 575–586.

Murray AM (2001) The fossil record and biogeography of the Cichlidae (Actinopterygii: Labroidei). *Biological Journal of the Linnean Society*, **74**, 517–532.

Murray AM, Stewart KM (1999) A new species of tilapiine cichlid from the Pliocene, Middle Awash, Ethiopia. *Journal of Vertebrate Paleontology*, **19**, 293–301.

Muruga B, Anyango B (2013) A survey of extremophilic bacteria in Lake Magadi, Kenya. *American Journal of Molecular and Cellular Biology*, 14–26.

Muruga BN (2013) Isolation Of bacteriophage infecting haloalkaliphilic bacteria In Lake Magadi, Kenya. *International of Journal of Innovative Research and Development*, **2**, 257–261.

Muruga BN, Wagacha JM, Kabaru JM, Amugune N (2014) Effect of physicochemical conditions on growth rates of cyanobacteria species isolated from Lake Magadi, a soda lake in Kenya. *WebPub Journal of Scientific Research*, **2**, 53–62.

Nagl S, Tichy H, Mayer WE *et al.* (2000) The origin and age of haplochromine fishes in Lake Victoria , East Africa. *Proceedings of the Royal Society B: Biological Sciences*, **267**, 1049–1061.

Nagl S, Tichy H, Mayer WE *et al.* (2001) Classification and phylogenetic relationships of African tilapiine fishes inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **20**, 361–374.

Narahara A, Bergman HL, Laurent P *et al.* (1996) Respiratory physiology of the Lake Magadi tilapia (*Oreochromis alcalicus grahami*), a fish adapted to a hot, alkaline, and frequently hypoxic environment. *Physiological Zoology*, **69**, 1114–1136.

Nichols JT (1923) A new wrasse and two new cichlids from northeast Africa. *American Museum Novitates*, **65**, 1–4.

Norman JR (1922) Four new fishes from Tanganyika Territory. *Annals and Magazine of Natural History (Series 9)*, **9**, 686–688.

Nosil P (2012) *Ecological speciation*. Oxford University Press, Oxford.

Nosil P, Harmon LJ, Seehausen O (2009) Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, **24**, 145–156.

Olaka LA, Odada EO, Trauth MH, Olago DO (2010) The sensitivity of East African rift lakes to climate fluctuations. *Journal of Paleolimnology*, **44**, 629–644.

Osborn HF (1902) The law of adaptive radiation. *The American Naturalist*, **36**, 353–363.

Palaiokostas C, Bekaert MEL, Khan MGQ *et al.* (2013) Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) using RAD sequencing. *PLoS ONE*, **8**, e68389.

Papah MB, Kisia SM, Ojoo RO *et al.* (2013) Morphological evaluation of spermatogenesis in Lake Magadi tilapia (*Alcolapia grahami*): A fish living on the edge. *Tissue and Cell*, **45**, 371–382

Parchman TL, Benkman CW, Britch SC (2006) Patterns of genetic variation in the adaptive radiation of New World crossbills (Aves: Loxia). *Molecular Ecology*, **15**, 1873–1887.

Peters WCH (1852) Diagnosen von neuen Flussfischen aus Mossambique. *Bericht über die zur Bekanntmachung geeigneten Verhandlungen der Königlichen Preussischen Akademie der Wissenschaften zu Berlin*, **1**, 681–685.

Petren K, Grant PR, Grant BR, Keller LF (2005) Comparative landscape genetics and the adaptive radiation of Darwin's finches: The role of peripheral isolation. *Molecular Ecology*, **14**, 2943–2957.

Poletto AB, Ferreira IA, Cabral-de-Mello D *et al.* (2010) Chromosome differentiation patterns during cichlid fish evolution. *BMC Genetics*, **11**, 50.

Poll M (1948) Descriptions de Cichlidae nouveaux recueillis par la mission hydrobiologique belge au Lac Tanganika (1946-1947). *Bulletin du Musée Royal d'Histoire Naturelle de Belgique*, **24**, 1–31.

Randall D, Wood CM, Perry SF *et al.* (1989) Urea excretion as a strategy for survival in a fish living in a very alkaline environment. *Nature*, **337**, 165–166.

Reardon EE, Chapman LJ (2010) Hypoxia and energetics of mouth brooding: is parental care a costly affair? *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **156**, 400–406.

Recknagel H, Elmer KR, Meyer A (2013a) A hybrid genetic linkage map of two ecologically and morphologically divergent Midas cichlid fshes (*Amphilophus* spp.) obtained by massively parallel DNA sequencing (ddRADSeq). *G3*, **3**, 65–74.

Recknagel H, Kusche H, Elmer KR, Meyer A (2013b) Two new endemic species in the Midas cichlid complex from Nicaraguan crater lakes: *Amphilophus tolteca* and *Amphilophus viridis* (Perciformes, Cichlidae). *Aqua, International Journal of Ichthyology*, **19**, 4–25.

Reite OB, Maloiy GMO, Aasehaug B (1974) pH, salinity and temperature tolerance of Lake Magadi *Tilapia*. *Nature*, **247**, 315.

Relkin R (1989) Pineal response in the cichlid fish *Oreochromis mossambicus* to extreme osmolality, pH, and temperature. *Journal of Pineal Research*, **7**, 37–43.

Rico C, Turner GF (2002) Extreme microallopatric divergence in a cichlid species from Lake Malawi. *Molecular Ecology*, **11**, 1585–1590.

Riedel R, Costa-Pierce BA (2005) Feeding ecology of Salton Sea Tilapia (*Oreochromis* spp). *Bulletin of the Southern California Academy of Sciences*, **104**, 26–36.

Rivas LR (1964) A reinterpretation of the concepts "sympatric" and "allopatric" with proposal of the additional terms "syntopic" and “allotopic.” *Systematic Zoology*, **13**, 42–43.

Roberts N, Taieb M, Barker P *et al.* (1993) Timing of the Younger Dryas event in East Africa from lake level changes. *Nature*, **366**, 146–148.

Robichaux RH, Carr GD, Liebman M, Pearcy RW (1990) Adaptive radiation of the Hawaiian Silversword Alliance (Compositae - Madiinae): Ecological, morphological and physiological diversity. *Annals of the Missouri Botanical Garden*, **77**, 64–72.

Rüppell E (1852) *Verz Mus Senkenberg naturforsch Ges aufges Sammlungen*. Frankfurt-am-Main, Frankfurt.

Salzburger W, Meyer A (2004) The species flocks of East African cichlid fishes: recent advances in molecular phylogenetics and population genetics. *Naturwissenschaften*, **91**, 14.

Salzburger W, Mack T, Verheyen E, Meyer A (2005) Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. *BMC Evolutionary Biology*, **5**, 17.

Salzburger W, Van Bocxlaer B, Cohen AS (2014) Ecology and evolution of the African Great Lakes and their faunas. *Annual Review of Ecology, Evolution, and Systematics*, **45**, 519–545.

Samonte IE, Satta Y, Sato A *et al.* (2007) Gene flow between species of Lake Victoria haplochromine fishes. *Molecular Biology and Evolution*, **24**, 2069–2080.

Sari M (2008) Threatened fishes of the world: *Chalcalburnus tarichi* (Pallas 1811) (Cyprinidae) living in the highly alkaline Lake Van, Turkey. *Environmental Biology of Fishes*, **81**, 21–23.

Sauvage H-E (1884) Note sur des poissons de Franceville, Haute Ogooué. *Bulletin de la Société Zoologique de France*, **9**, 193–198.

Schliewen UK, Rassmann K, Markmann M *et al*. (2001) Genetic and ecological divergence of a monophyletic cichlid species pair under fully sympatric conditions in Lake Ejagham, Cameroon. *Molecular Ecology*, **10**, 1471–1488.

Schliewen UK, Tautz D, Paabo S (1994) Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature*, **368**, 629–632.

Schluter D (1993) Adaptive radiation in sticklebacks: Size, shape and habitat use efficiency. *Ecology*, **74**, 699–709.

Schluter D (2000) *The ecology of adaptive radiation*. Oxford University Press, Oxford.

Schluter D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution*, **16**, 372–380.

Schluter D, Mcphail JD (1993) Character displacement and replicate adaptive radiation. **8**, 197–200.

Schluter T, Kohring R, Mehl J (1992) Hyperostotic fish bones ("Tilly bones") from presumably Pliocene phosphorites of the Lake Manyara area, northern Tanzania. *Paläontologische Zeitschrift*, **66**, 129–136.

Schwan TG, Lamberti GA (1986) Influence of oxygen concentration on the respiratory behaviour of tilapia (*Sarotherodon alcalicus grahami*) in Lake Nakuru, Kenya. *African Journal of Ecology*, **24**, 199–202.

Schwarzer J, Misof B, Tautz D, Schliewen U (2009) The root of the East African cichlid radiations. *BMC Evolutionary Biology*, **9**, 186.

Seegers L (2008) The fishes collected by G A Fischer in East Africa in 1883 and 1885/86. *Zoosystematics and Evolution*, **84**, 149–195.

Seegers L, Tichy H (1999) The *Oreochromis alcalicus* flock (Teleostei: Cichlidae) from Lake Natron and Magadi, Tanzania and Kenya, with description of two new species. *Ichthyological Explorations of Freshwaters*, **10**, 97–146.

Seegers L, De Vos L, Okeyo DO (2003) Annotated checklist of the freshwater fishes of Kenya (excluding the lacustrine haplochromines from Lake Victoria). *Journal of East African Natural History*, **92**, 11–47.

Seegers L, Sonnenberg R, Tichy H (2001) The *Alcolapia* group, a remarkable species flock from lakes Natron, Tanzania, and Magadi, Kenya: a further piece of the puzzle of cichlid evolution? *Journal of Aquariculture and Aquatic Sciences*, **9**, 335–364.

Seegers L, Sonnenberg R, Yamamoto R (1999) Molecular analysis of the *Alcolapia* flock from lakes Natron and Magadi, Tanzania and Kenya (Teleostei: Cichlidae), and implications for their systematics and evolution. *Ichthyological Explorations of Freshwaters*, **10**, 175–199.

Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, **19**, 198–207.

Seehausen O (2006) African cichlid fish: A model system in adaptive radiation research. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 1987–1998.

Seehausen O (2007) Chance, historical contingency and ecological determinism jointly determine the rate of adaptive radiation. *Heredity*, **99**, 361–363.

Seehausen O (2015) Process and pattern in cichlid radiations - inferences for understanding unusually high rates of evolutionary diversification. *New Phytologist*, doi: 10.1111–nph.13450.

Seyoum S, Kornfield I (1992) Taxonomic notes on the *Oreochromis niloticus* subspecies-complex (Pisces: Cichlidae), with a description of a new subspecies. *Canadian Journal of Zoology*, **70**, 2161–2165.

Simpson GG (1953) *The major features of evolution*. Columbia University Press, New York.

Snorrason SS, Skulason S (2004) Adaptive speciation in northern freshwater fish—patterns and processes. In: *Adaptive Speciation* (Eds. Dieckmann U, Doebeli M, Metz JAJ, Tautz D). Cambridge University Press, Cambridge.

Soulebeau A, Aubriot X, Gaudeul M *et al.* (2015) The hypothesis of adaptive radiation in evolutionary biology: hard facts about a hazy concept. *Organisms Diversity and Evolution*, doi: 10.1007/s13127-015-0220-z.

Soussa H, Azab A, Kashaigili JJ *et al.* (2010) *Hydrological and environmental aspects of wetlands: Analytical tools for wetlands management of the Nile basin*. Nile Basin Capacity Building Network, Cairo.

Stauffer JR, Van Snik Gray E (2004) Phenotypic plasticity: its role in trophic radiation and explosive speciation in cichlids (Teleostei: Cichlidae). *Animal Biology*, **54**, 137–158.

Steindachner F (1864) Ichthyologische Mittheilungen (VII). *Verhandlungen der K-K zoologisch-botanischen Gesellschaft in Wien*, **14**, 223–232.

Stelkens R, Seehausen O (2009) Phenotypic divergence but not genetic distance predicts assortative mating among species of a cichlid fish radiation. *Journal of Evolutionary Biology*, **22**, 1679–1694.

Tata Chemicals Ltd (2011) *Tata Magadi Report*. Tata Chemicals Magadi.

Tata Chemicals Ltd. *Magadi Facility* webpage. Available at: www.tatachemicals.com/magadi/products/overview.html Accessed: July 2015

Tebbs EJ, Remedios JJ, Avery ST, Harper DM (2013) Remote sensing the hydrological variability of Tanzania's Lake Natron, a vital Lesser Flamingo breeding site under threat. *Ecohydrology & Hydrobiology*, **13**, 148–158.

Thys van den Audenaerde DFE (1964) Révision systématique des espèces congolaises du genre Tilapia (Pisces, Cichlidae). *Annales, Musée Royal de l'Afrique Centrale, Tervuren, Série in 8o, Sciences Zoologiques*, **124**, 1–155.

Thys van den Audenaerde DFE (1968) *An annotated bibliography of Tilapia (Pisces, Cichlidae)*. Royal Museum Central Africa, Tervuren.

Tichy H, Seegers L (1999) The *Oreochromis alcalicus* flock (Teleostei Cichlidae) from lakes Natron and Magadi, Tanzania and Kenya: A model for the evolution of new species flocks in historical times. *Ichthyological Explorations of Freshwaters*, **10**, 147–174.

Trewavas E (1933) Scientific results of the Cambridge expedition to the East African Lakes, 1930-1. The cichlid fishes. *Zoological Journal of the Linnean Society*, **38**, 309–341.

Trewavas E (1941) Nyasa fishes of the genus *Tilapia* and a new species from Portuguese East Africa. *Journal of Natural History Series 11*, **7**, 294–306.

Trewavas E (1966a) Fishes of the genus *Tilapia* with four anal spines in Malawi, Rhodesia, Mozambique and southern Tanzania. *Revue de Zoologie et de Botanique Africaines*, **74**, 50–62.

Trewavas E (1966b) A preliminary review of fishes of the genus *Tilapia* in the eastward-flowing rivers of Africa, with proposals of two new specific names. *Revue de Zoologie et de Botanique Africaines*, **74**, 394–424.

Trewavas E (1973) A new species of cichlid fishes of rivers Quanza and Bengo, Angola, with a list of the known Cichlidae of these rivers and a note on *Pseudocrenilabrus natalensis* Fowler. *Bulletin of the British Museum (Natural History) Zoology*, **25**, 27–37.

Trewavas E (1982a) Generic groupings of *Tilapiini* used in aquaculture. *Aquaculture Research*, **27**, 79–81.

Trewavas E (1982b) Tilapias: taxonomy and speciation. In: *The Biology and Culture of Tilapias* (eds Pullin R, Lowe-McConnel RH), pp. 3–13. ICLARM Conference Proceedings, Manila, Philippines.

Trewavas E (1983) *Tilapiine Fishes of the genera Sarotherodon, Oreochromis and Danakilia*. British Museum (Natural History), London.

Van Couvering J (1982) Fossil cichlid fish of Africa. In: *Special Papers in Palaeontology*, pp. 29: 1–103. The Palaeontological Association, London.

Vanden Bossche J-P, Bernacsek GM (1990) *Source book for the inland fishery resources of Africa 1. CIFA Technical Paper. No. 18.1.* Food and Agriculture Organisation, Rome.

Vareschi E (1979) The ecology of Lake Nakuru (Kenya). II. Biomass and spatial distribution of fish (*Tilapia grahami* Boulenger = *Sarotherodon alcalicum grahami* Boulenger). *Oecologia*, **37**, 321–335.

Verheyen E, Salzburger W, Snoeks J, Meyer A (2003) Origin of the superflock of cichlid fishes from Lake Victoria, East Africa. *Science*, **300**, 325–329.

Via S (2011) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 451–460.

Wagner CE, Harmon LJ, Seehausen O (2014) Cichlid species-area relationships are shaped by adaptive radiations that scale with area. *Ecology Letters*, **17**, 583–592.

Wagner CE, Keller I, Wittwer S *et al.* (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**, 787–798.

Walsh PJ, Grosell M, Goss GG *et al.* (2001) Physiological and molecular characterization of urea transport by the gills of the Lake Magadi tilapia (*Alcolapia grahami*). *Journal of Experimental Biology*, **204**, 509–520.

Wellenreuther M, Sánchez-Guillén RA (2015) Non-adaptive radiation in damselflies. *Evolutionary Applications*, **1**, 1–16.

White TH (1953) Some speculations on the sudden occurrence of floods in the history of Lake Magadi. *Journal of The East Africa Natural History Society*, **22**, 69–71.

Wilkie MP, Wood CM (1996) The adaptations of fish to extremely alkaline environments. *Comparative Biochemistry And Physiology Part B: Biochemistry and Molecular Biology*, **113**, 665–673.

Williamson D, Taieb M, Damnati B *et al.* (1993) Equatorial extension of the Younger Dryas event: rock magnetic evidence from Lake Magadi (Kenya). *Global and Planetary Change*, **7**, 235–242.

Wilson PJ, Wood CM, Walsh PJ *et al.* (2004) Discordance between genetic structure and morphological, ecological, and physiological adaptation in Lake Magadi tilapia. *Physiological and Biochemical Zoology*, **77**, 537–555.

Won Y-J, Wang Y, Sivasundar A *et al*. (2006) Nuclear gene variation and molecular dating of the cichlid species flock of Lake Malawi. *Molecular Biology and Evolution*, **23**, 828–837.

Wood CM, Bergman HL, Bianchini A *et al.* (2011) Transepithelial potential in the Magadi tilapia, a fish living in extreme alkalinity. *Journal of Comparative Physiology B*.

Wood CM, Bergman HL, Laurent P *et al.* (1994) Urea production, acid-base regulation and their interactions in the Lake Magadi tilapia, a unique teleost adapted to a highly alkaline environment. *Journal of Experimental Biology*, **189**, 13–36.

Wood CM, Wilson P, Bergman HL *et al.* (2002a) Obligatory urea production and the cost of living in the Magadi tilapia revealed by acclimation to reduced salinity and alkalinity. *Physiological and Biochemical Zoology*, **75**, 111–122.

Wood CM, Wilson P, Bergman HL *et al.* (2002b) Ionoregulatory strategies and the role of urea in the Magadi tilapia (*Alcolapia grahami*). *Canadian Journal of Zoology*, **80**, 503–515.

Woodhouse GW (1912) The fish in Lake Magadi. *The Journal of the East Africa and Uganda Natural History Society*, **2**, 95–97.

Wright PA, Perry SF, Randall D *et al*. (1990) The effects of reducing water pH and total CO_2 on a teleost fish adapted to an extremely alkaline environment. *Journal of Experimental Biology*, **151**, 361–369.

Zaccara S, Crosa G, Vanetti I *et al.* (2014) Genetic and morphological analyses indicate high population mixing in the endangered cichlid *Alcolapia* flock of East Africa. *Conservation Genetics*, **15**, 429–440.

Zavarzina DG, Zhilina TN, Kuznetsov BB *et al.* (2013) *Natranaerobaculum magadiense* gen. nov., sp. nov., a new anaerobic, alkalithermophilic bacterium from soda lake Magadi. *International Journal of Systematic and Evolutionary Microbiology*, **63**, 4456–4461.

Chapter two

Methods overview

Abstract

Integrative studies are increasingly being used to disentangle longstanding and complex questions in evolutionary biology. Here, the general methods employed in the analysis sections for this thesis are introduced and provide a basis for the data chapters three-six. This chapter discusses the merits, benefits and potential limitations of these methods, also outlining what may be gained from the integration of varied techniques in the scientific process. The exact protocols and parameters employed for each method in the present study are not included here, but given in detail in the methods section of each respective chapter.

Aims

This chapter aims to introduce the general methods used in this thesis, and discuss the benefits and limitations of the methods for their respective applications, explaining the choice of methods to address the questions at hand.

RAD sequencing – Chapters three and four

The advent of next generation sequencing (NGS) technologies has revolutionised the fields of phylogenetics, phylogeography and population genetics, allowing unprecedented levels of sequence data to be generated. A single NGS dataset can be used for multiple analyses and studies, rather than the traditional Sanger sequencing methods employing different markers dependent on downstream analysis (i.e., concatenated mtDNA and nuclear markers for phylogenetics vs. AFLPs and microsatellites for population genetics; Brito & Edwards 2008). Furthermore, such massively parallel sequencing technologies are cheaper (per base sequenced) and quicker than earlier methodologies (Liu *et al.* 2012). The rapid uptake and application of these sequencing methodologies to phylogenetic and population questions is demonstrated in the substantial increase in publications

based on such methods, relative to the near-constancy in publications using Sanger sequencing (Figure 2.1).

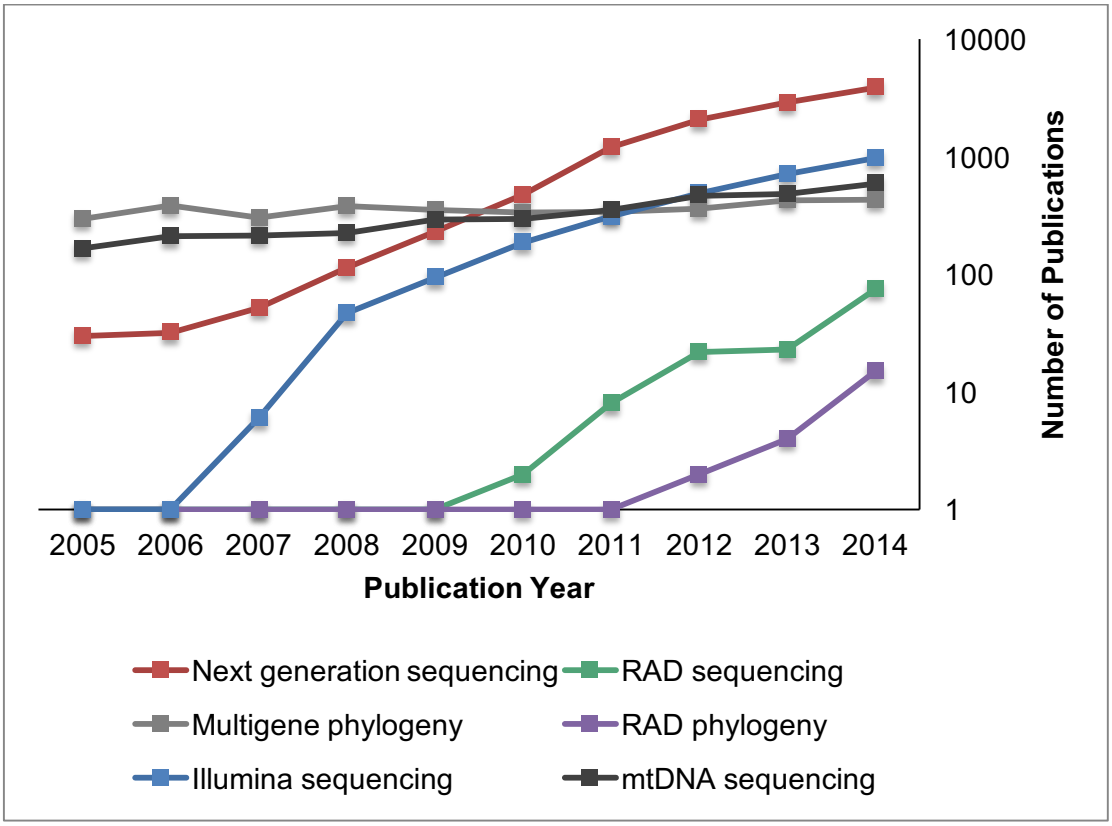


Figure 2.1. The increasing use of next generation sequencing technologies in scientific publications. Searches conducted on NCBI Pubmed using the search terms as indicated in the figure legend (all conducted March 2015). The annual number of publications returned from search terms associated with Sanger sequencing or traditional phylogenetic methods (grey) showed only slight increases over the time period 2005-2014, while those terms associated with next generation sequencing (colour) showed substantial increases in publication numbers.

NGS technologies have application across a wide range of fields, including complex trait mapping (Day-Williams & Zeggini 2010); adaptation (Stapley *et al.* 2010); and ecological genetics (Davey *et al.* 2010; 2011).

Although NGS technologies now allow for sequencing of large amounts of DNA rapidly and cost-effectively, sequencing entire genomes for species and population comparison is still often technically and financially prohibitive for eukaryotic genomes of gigabases in size. Thus, a compromise of more detailed sequence data detailing individual single nucleotide polymorphisms (SNPs), but at a reduced representation than the entire genome allows more individuals to be included in the analysis. Surveying genome-wide sequence variation at varying unlinked loci can

provide a suitable alternative. Such reduced representation methods have been developed for population level analyses, including the use of restriction site-associated DNA (RAD) sequencing (Baird *et al.* 2008; Emerson *et al.* 2010; Hohenlohe *et al.* 2010).

Restriction enzymes and their corresponding restriction endonuclease recognition sites have for a long time been harnessed to analyse polymorphisms and SNPs across populations, perhaps most significantly in the form of anonymous loci methods such as restriction fragment length polymorphisms (RFLPs; Botstein *et al.* 1980) and amplified fragment length polymorphisms (AFLPs; Vos *et al.* 1995). Sequencing of the genome around the restriction enzyme recognition sites offers the potential to analyse these polymorphisms in a non-anonymous manner. Restriction-site associated DNA (RAD) sequencing was first developed as a method to provide exactly this sequence coverage, and worked by hybridising sequences ('tags') surrounding each restriction site to DNA microarrays (Miller *et al.* 2007a; b; Lewis *et al.* 2007). However, the use of microarrays required either the use of existing genomic resources in model organisms such as *Drosophila* or the development of an assay-specific microarray requiring cloning and pre-selection of informative markers (Miller *et al.* 2007b). The expansion of this protocol to utilise NGS platforms in sequencing the RAD tags directly rather than hybridising to microarrays greatly enhanced the accessibility and application to non-model organisms, as well as the rapid discovery of SNPs and high throughput of large population samples (Baird *et al.* 2008). The method can be modified to select for number/density of loci through use of specific restriction enzymes (mainly in the choice of frequent or rare cutters), and increased number of restriction enzymes used. Formal modifications to the original RAD protocol have included the use of a double digestion with a combined rare and frequent cutter enzyme followed by size selection (ddRAD, Peterson *et al.* 2012; ezRAD, Toonen *et al.* 2013), and the use of type IIB restriction endonucleases to constrain tag length (2B-RAD, Wang *et al.* 2012).

Using restriction sites to reduce representation targets orthologous regions as restriction sites are conserved across species (Rubin *et al.* 2012). RAD provides greater information than anonymous loci approaches (such as AFLPs) and can be mapped to a reference genome where available. In addition, RAD provides the benefit of identifying novel SNPs without the requirement for screening novel genetic markers for distinction between species (Davey *et al.* 2010), and produces a dense coverage (>10,000) of SNPs. The use of barcoded tags denoting specimen

ID attached to each sheared fragment to identify samples amongst pooled libraries allows for fragments from several individuals to be multiplexed and run together.

RAD protocol and data processing

The protocol for preparing RAD sequencing libraries is summarised in Figure 2.2, but briefly comprises digestion of genomic DNA with restriction enzyme; ligation of P1 adapter sequence; pooling of samples; random shearing to average size of 500 bp; gel extraction of fragments of size 300-700 bp; blunting of DNA ends and addition of adenine overhangs; ligation of P2 adapter; followed by final PCR amplification and gel purification (Baird *et al.* 2008). Paired-end sequencing (i.e., sequencing from both ends of the read) increases coverage and sequence depth around the restriction site. As the size of fragments range across a pre-determined size range (decided by the gel-extraction selection steps; typically 300-700 bp), and a short sequence (100-200 bp) at each end of each fragment is sequenced, the combined reads can provide coverage for a sequence length of up to 1400 bp around each restriction site, although sequencing depth will vary across this extent.

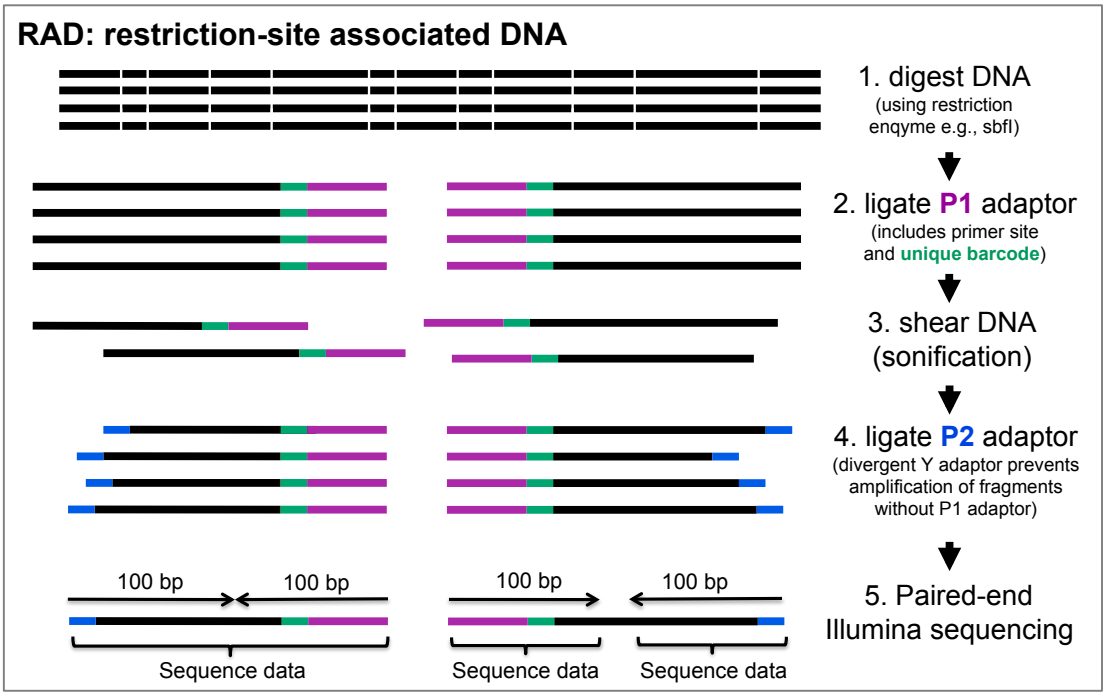


Figure 2.2 RAD sequencing protocol.

Modified from Baird *et al.* 2008. DNA is restricted, typically using a rare (8-base) cutter such as *Sbf*I. Specific enzymes used depend on application and may employ a more frequent cutter (e.g., 6-base recognition site, such as *EcoR*I) if more fragments are required, or use two enzymes together to fine-tune sizing of fragments or avoid use of shearing (e.g., ddRAD protocol). Shearing and size selection produces fragments of different sizes, such that paired end sequencing results in all fragments being sequenced at the restriction site end, but the sheared end will vary in sequence position.

Figure 2.3 depicts the build-up of sequence coverage from the overlaying of sequencing reads around a restriction site, visualising data from RAD sequencing performed in the current thesis and aligned to the *O. niloticus* reference genome – methods as detailed in chapter three. The total sequencing extent around each restriction site is termed a RAD-tag (locus), and can be used downstream in a similar manner to Sanger-sequenced markers. More typically, the SNP information is extracted from each locus and used in a dataset for population analyses (although the full matrix of sequence data including polymorphic and invariant sites is typically more suitable for phylogenetic analysis – see discussion in chapter three). Often, inclusion will be limited to one SNP per locus to prevent the inclusion of very tightly linked markers – although further separation may be necessary to completely address the issue of linkage disequilibrium in the use of these markers. Where the reference genome is available and thus position of sequenced regions can be attributed, the extent of linkage disequilibrium can be tested to exclude linked sites

in downstream analysis (e.g., Martin *et al.* 2013; Ford *et al.* 2015; see also discussion in chapters three and four). While such SNP data is employed in a wide range of population genetic and genome differentiation tests, full sequence data (i.e., including invariant sites) is typically more suitable for phylogenomic purposes, providing more accurate reconstruction for branch placement (Wagner *et al.* 2013; Ford *et al.* 2015) and branch length (Lemmon & Lemmon 2013; Martin *et al.* 2015).

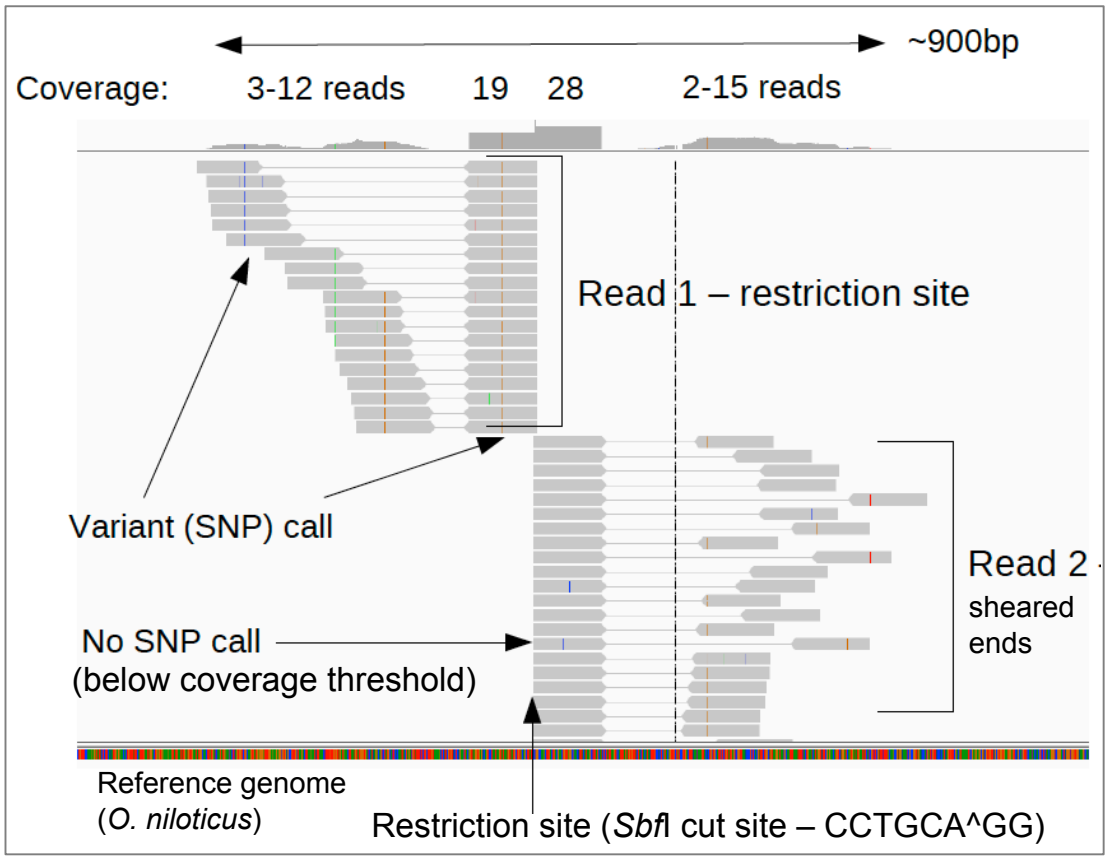


Figure 2.3. Visualisation of a RAD-Seq tag. Paired-end sequenced reads aligned to a reference genome from the current analysis (chapter three). The shearing step in the RAD protocol means that fragments of differing lengths are sequenced around the restriction site. The ends closest to the restriction site is sequenced for every read, providing high coverage, while the location of the fragmented end sequenced differs for each read, providing lower coverage, but greater extent around the site, resulting in sequencing of ~1000bp per locus, 500bp each side of the restriction site.

The processing of NGS generated data requires various protocols in the stringency of quality control and filtering, as sequencing technology is less accurate than Sanger sequencing (Fox *et al.* 2014) and the large amounts of data produced mean it is unfeasible to check sequence reads or alignments by eye. Furthermore, no visual output is produced (such as chromatograms in the case of Sanger

sequencing), so rules must be applied in the assessment of sufficient data or coverage to accurately call a base, and for cases of heterozygosity vs. ambiguity. Several recent analyses have advocated either the necessity for maximal stringency (Henning *et al.* 2014) or cautioned the loss of informative data (particularly in the form of outlier loci) in removing high proportions of data at the quality control stages (Huang & Knowles 2014).

RAD sequencing is subject to a number of biases, including: restriction fragment length bias, amplification bias associated with GC content, and overrepresentation of regions with high GC content (DaCosta & Sorenson 2014). As a result, the read depths are highly variable at each RAD locus, meaning that alleles cannot be separated by copy number using read depth. Heterozygosity of restriction sites (the premise on which RFLP and AFLP methods are based), may cause problems in analysis as software designed for whole genome analysis does not recognise heterozygosity of presence/absence, and heterozygous restriction sites will also influence read depth of adjacent sites due to fragment length bias (Davey *et al.* 2012). While exclusion of all sites for which there are missing data would address these biases, it is often not necessary and alignment to a reference genome generally means that genotypes can be accurately called even at sites with missing genotypes (Davey *et al.* 2012). Furthermore, although (as with any NGS protocol) RAD sequencing requires good quality DNA, it has been shown to be robust to moderately degraded DNA (Graham *et al.* 2015).

Application of RAD sequencing to investigate cichlid phylogeography

Recent cichlid radiations are notoriously difficult to resolve phylogenetically using traditional sequencing techniques, which has been attributed to incomplete lineage sorting in the recent separation time of young species (e.g., Wagner *et al*. 2013). Even large-scale multi-gene phylogenies have failed to resolve species in closely related species flocks (e.g., Won *et al.* 2006), and frequently population genetics or phylogeographic techniques are required to delimit the close species, typically using microsatellites (Wagner *et al.* 2012) or AFLPs (Genner & Turner 2012) that fare better in species diagnosis for very recently diverged groups. While these techniques may successfully differentiate lineages, they do not provide lasting genomic (sequence) information and, being reliant on subjective assessment of allele or peak size, are not readily comparable between studies. Although initially developed for population analyses, RAD methodology has been tested for species distinction (Rubin *et al.* 2012) and recent RAD projects have used the method for

cichlid phylogenies, achieving resolution at an unprecedented level for the Lake Victoria adaptive radiation (Wagner *et al.* 2013). As such, a genomic approach using RAD to define the evolutionary relationships of *Alcolapia* is likely to yield greater resolution than possible from multigene dataset based on traditional sequencing of individual genes. In addition, species delimitation is difficult in recent radiations using genealogical approaches, and analysis suggests that population genomic approaches based on large sets of SNPs are more reliable in delimiting recently derived species (Shaffer & Thomson 2007).

Such powerful genomic approaches may prove more useful for investigating the species status of *Alcolapia* as genetic approaches based on mtDNA have so far failed to resolve the constituent species (Seegers *et al.* 1999; Wilson *et al.* 2004). Furthermore, it is possible that the actual diversity is greater than the four described species since these fish exhibit considerable phenotypic variability (Tichy & Seegers 1999). Preliminary analysis of mtDNA for the present study was also unable to differentiate species or populations within the soda lake radiation (Figure 2.4, methods as described in Appendix 2.1), and even the most variable region of the mitochondrial DNA (control region) had insufficient segregating sites to separate the species. Thus, the application of increased amounts of sequence data may provide sufficient information to phylogenomically differentiate these species.

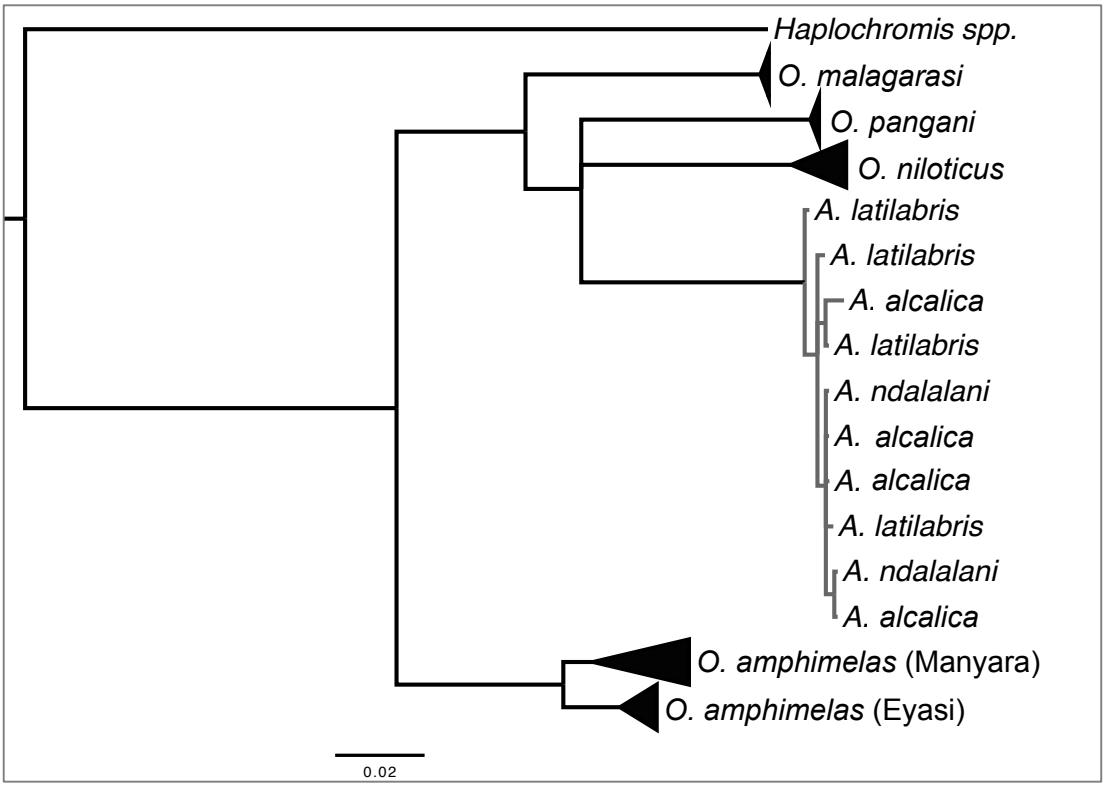


Figure 2.4. Phylogenetic tree of *Alcolapia* (mtDNA control region).

The Bayesian inference tree includes *Oreochromis* and *Alcolapia* species, rooted using *Haplochromis* (H. spp). BPP values are >0.85 for all nodes in black. There is no resolution of or branch support for *Alcolapia* species or populations using this molecular marker. Methods for the sequencing and phylogenetic analysis are described in Appendix 2.1.

Trophic niche segregation: Stable isotope analysis and stomach contents analysis – Chapters five and six

Stable isotope analysis (SIA) is a useful tool for examining diet composition based on specific isotopic signatures acquired by consumers from different food sources (DeNiro & Epstein 1978; 1981). Isotopic ratios of several elements have been employed to estimate trophic partitioning and food web position. Isotopes of a given element differ in atomic mass (based on the number of neutrons contained), and as stable isotopes (nonradiogenic isotopes) do not decay, the relative ratio of the heavy to light isotope can be used to track resources through a food system (Newton 2010). These ratios are quantified using mass spectrometry, and results given as the comparison to an international standard using delta notation (McKinney *et al.* 1950) where:

$$\delta(\text{isotope}) = \text{Ratio(sample)}/\text{Ratio(standard)} - 1$$

and the δ value is given for the heavy isotope of a pair (Newton 2010), and values are multiplied by 1000 and given in units of per mille (‰)**.** Carbon and nitrogen are the most commonly used for ecological purposes, but additional organic elements have been used when carbon and nitrogen together are incapable of separating sources (e.g., Neill & Cornwell 1992). Other elements may also have better precision in particular situations, for example sulphur has a higher mean difference between producers in the marine environment (Connolly *et al.* 2004), and is capable of distinguishing between pelagic and benthic sources, while hydrogen demonstrates accuracy in discriminating aquatic from terrestrial (or emergent) vegetation (Middelburg 2014). However, carbon and nitrogen remain the most commonly used isotopes in ecological studies. Carbon and nitrogen both have two isotopes: C^{12} (common) and C^{13} (rare); N^{14} (common) and N^{15} (rare) (Newton 2010).

Isotope ratios provide additional information to stomach contents analysis, as the ratios represent dietary intake over a longer time period and indicate resources assimilated rather than only those ingested (Vander Zanden *et al.* 1999; Marijnissen *et al.* 2008). The isotopic signature of the prey items becomes assimilated into the consumer's tissues, typically over the scale of weeks to months in fish muscle tissue (Boecklen *et al.* 2011; Weidel *et al.* 2011), although this appears to be dependent on body size (Vander Zanden *et al.* 2015). The amount by which the ratios change between trophic levels is the fractionation value, depicted as Δ(isotope), and it is the differences in the pattern of these changes between nitrogen and carbon that make them useful for trophic food ecology. Nitrogen isotope ratios are used to identify the

trophic level of an organism, which are typically enriched by 2-4‰ relative to its food source (DeNiro & Epstein 1981; Vander Zanden *et al.* 1999; Post 2002; Vanderklift & Ponsard 2003; Caut *et al.* 2009). Because nitrogen ratios tend to have a large and constant discrimination factor between trophic levels, it can be used as a proxy for trophic niche, although this only remains relative to other groups/samples tested, and absolute trophic level position is dependent on ascertaining the nitrogen ratio of relevant food sources (or, particularly, those at the base of the specific food chain). The carbon ratios show a much smaller discrimination factor between each trophic level (DeNiro & Epstein 1978; Post 2002) and are therefore typically ignored as negligible with the assumption that carbon ratios will almost exactly mirror those of the sources consumed. Where mixed resources are consumed, the consumer signal will reflect a mixed ratio relative to the proportions of each resource consumers (Middelburg 2014). As such, while consumers feeding on a single source will typically reflect directly the carbon source exactly (and the nitrogen source plus the discrimination factor), those feeding on multiple resources will appear at midpoints. Thus, primary consumers may vary along carbon ratios reflecting mixed diets, and omnivores feeding at multiple trophic levels will reflect intermediate nitrogen levels, such that the representation of trophic level will be continuous rather than categorically based on the exact input sources to the system. This interplay of carbon and nitrogen values is commonly visualised as a biplot, which allows characterisation of isotopic niche space (Figure 2.5).

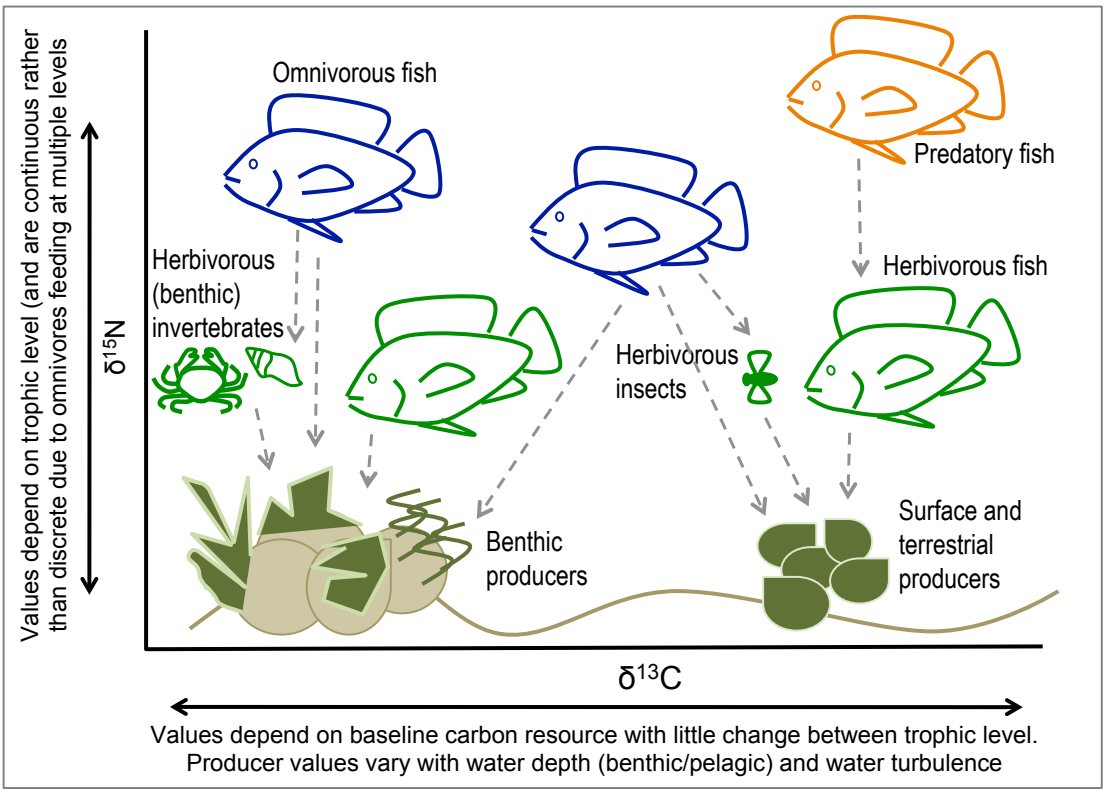


Figure 2.5. Schematic stable isotope bi-plot of $\delta^{13}C$ and $\delta^{15}N$, demonstrating how the use of the two isotopes can distinguish consumers based on resource type and trophic level. Grey arrows indicate resource use. Colours of fish depict trophic level: green, herbivorous; blue, omnivorous; orange, predatory (secondary consumer).

Carbon values of producers may overlap at baseline, meaning that carbon values higher up the food chain may not be sufficient to confidently assign initial food source. For example, in aquatic systems terrestrial and pelagic carbon sources often show a high degree of overlap in $\delta^{13}C$ (Pace *et al.* 2004), and $\delta^{13}C$ carbon values cannot reliably distinguish plants using C3, C4, or CAM photosynthetic pathways in aquatic systems (Keeley & Sandquist 1992), as plant values are more dependent on location in the water column and water turbulence (France 1995). Baseline $\delta^{13}C$ has been shown to vary with pH and concentration of dissolved CO_2 (Smyntek *et al.* 2012), indicating a depth profile variable that is not of concern in terrestrial food source analysis. Given this variance in transfer and input from baseline sources, observed values are highly dependent on system inputs, which vary over fine spatial scales, such that stable isotope ratios are not readily comparable between separate populations unless the baseline composition is known and can be accounted for (Middelburg 2014).

While SIA has typically been used for characterisation of food webs and describing feeding regimen shifts in single species, the characterisation of feeding

patterns (using stable isotope ratios as a proxy) can also be used to determine whether sympatric species exploit different trophic niches (Genner *et al.* 1999; Post 2002), and whether community-wide trophic metrics differ between sites (Layman *et al.* 2007). Such characterisation allows inferences to be made regarding the role of ecological speciation in adaptive radiation, as to whether species occurring sympatrically exhibit different diet specialisations from one another. Isotope analysis can also help to characterise population biogeography and connectivity (Newsome *et al.* 2007), which may be relevant in the Natron-Magadi system given the changing extent of population connectivity based on lake level fluctuations and flooding. SIA has previously demonstrated trophic niche segregation in Lake Malawi cichlids (Bootsma *et al.* 1996; Duponchelle *et al.* 2005; Anseeuw *et al.* 2009), although there are instances where dietary segregation does not fully explain sympatry (Genner *et al.* 1999). Additionally, SIA can be used to consider competition effects, since animal density may impact niche width and utilisation as a result of interspecific competition, whereby trophic divergence increases as areas become more heavily populated (with increased pressure on resources), varying with geographical location and season, (e.g., Bocher *et al.* 2000).

The complementarity of stable isotope analysis in combination with stomach contents analysis has previously been useful in separating cichlid species in larger food webs, for example where stomach contents can give an indication of feeding on different resources which may have the same isotope ratio, or where partitioning is spatial rather than resource-type based, e.g., benthic feeding indicated by sand presence in stomach (Duponchelle et al 2005). Furthermore, the methods complement each other by operating at different timescales. Previous gut content analysis in *A. grahami* has identified a diet comprising algal matter (including cyanobacteria) and invertebrates (copepods and dipterous larvae) (Coe 1966), although nothing is known about the diet of Lake Natron species. Given the volcanic source of the saline hot springs, there is likely to be a high degree of methanotrophy occurring at the primary producer level of the food web, which would be reflected in the stable isotope signatures of consumers further up the food chain. Methanotrophic bacteria can help fuel lake biomass, and consumer species previously thought to be strictly algal filter feeders have proven to be dependent on methanotrophic bacteria (Grey *et al.* 2004). Food sources based on methane have a lighter carbon isotope ratio (depleted ^{13}C) relative to other basal resources, and this pattern has successfully been used to distinguish diets of invertebrates by SIA, which showed differential uptake of methanotrophic and chemoautotrophic bacteria in lake chironomid larvae (Grey *et al.* 2004; Grey & Deines 2005).

Geometric morphometrics – Chapters five and six

The morphological study of fish (and other species) has largely been derived from the systematic study and definition of definable characters. As such, this approach of discrete characteristics (e.g., meristic counts of scales and fin rays) may provide useful information on characters defining morphological species, but be of less use in assessing overall body shape differentiation, or subtle changes between populations below the species level, and may be dependent on subjective assessment. Traditional morphometrics, which applies multivariate statistical methods to a set of variables (typically measurements between anatomical landmarks), allows examination of the covariation of shape with other variables of interest (Rohlf & Slice 1990). Such an approach allows examination of change in shape with size or environmental variables, but does not examine overall shape or relate changes in different measurements to each other. For example, the use of maximum width and length may be the same for several shapes of different outline, and measurements that rely on arbitrary maxima may not necessarily reflect homologous points with or those with biological relevance. The development of geometric morphometrics, which expanded the traditional morphometric approach to include geometric information about morphological structure using outline data or landmark points, is considered a paradigm shift in the field (Adams *et al.* 2002). The consideration of geometry and retention of information of the spatial covariation of landmarks was described at the time as 'a revolution in morphometrics' (Rohlf & Marcus 1993), and was accompanied by development of a rigorous statistical theory for shape analysis (Bookstein 1984; Kendall 1984; Bookstein 1991). The use of homologous landmarks across specimens rather than measurements allowed change in relative positioning to be analysed, while also allowing measurements between landmarks to be assessed (Rohlf & Marcus 1993). Furthermore, the correlation of landmarks allows statistical analysis to construct a consensus shape for the sample, and to reconstruct common ancestors between groups, as well as to visualise the shape differences responsible for variation between groups (Rohlf & Marcus 1993; Klingenberg 2013).

Geometric morphometrics is commonly used to quantify shape variation in development and ontogeny, systematics and phylogenetic history, response to environmental conditions, and effects of particular genes or genetic conditions (Klingenberg 2010). Furthermore, it has successfully been used to delimit species in cases of cryptic morphology where species were not separable by qualitative phenotypic data (e.g., Mutanen & Pretorius 2007; Francuski *et al.* 2009a; b; Gurgel-

Gonçalves *et al.* 2010) and in cases where genetic barcoding alone was insufficient to diagnose species (Chevasco *et al.* 2014; Schwarzfeld & Sperling 2014).

The process by which shape variation is examined in geometric morphometric analysis follows a standard protocol (Rohlf & Marcus 1993): landmark data is collected (2D or 3D); the relationship between landmark points is defined by a fitted function (typically thin plate spline; (Bookstein 1989); and the estimates of the fitted function are used in downstream multivariate statistical analyses (Figure 2.6). The thin-plate spline models rearrangement of landmarks as if placed on a metal sheet and the necessary deformation of the sheet to achieve relative movements of the landmarks to each other. This model thus allows the visualisation of additional shape changes (contour, outline) accompanying changes in landmark placement, even when outlines were not included in initial data collection, and these changes can be visualised as a warped grid. These methods are combined with relative warp analysis (Bookstein 1989), which is used to map landmarks onto a reference configuration (usually the mean configuration of all specimens in the analysis), to calculate deviations from the consensus to test directionality and statistical significance (Rohlf & Marcus 1993). Certainly, assessing shape change by way of deformed grids and warps seems more intuitive and straightforward in the investigation of biological relevance than assessing lists of coefficient values for multiple measurements in traditional morphometrics.

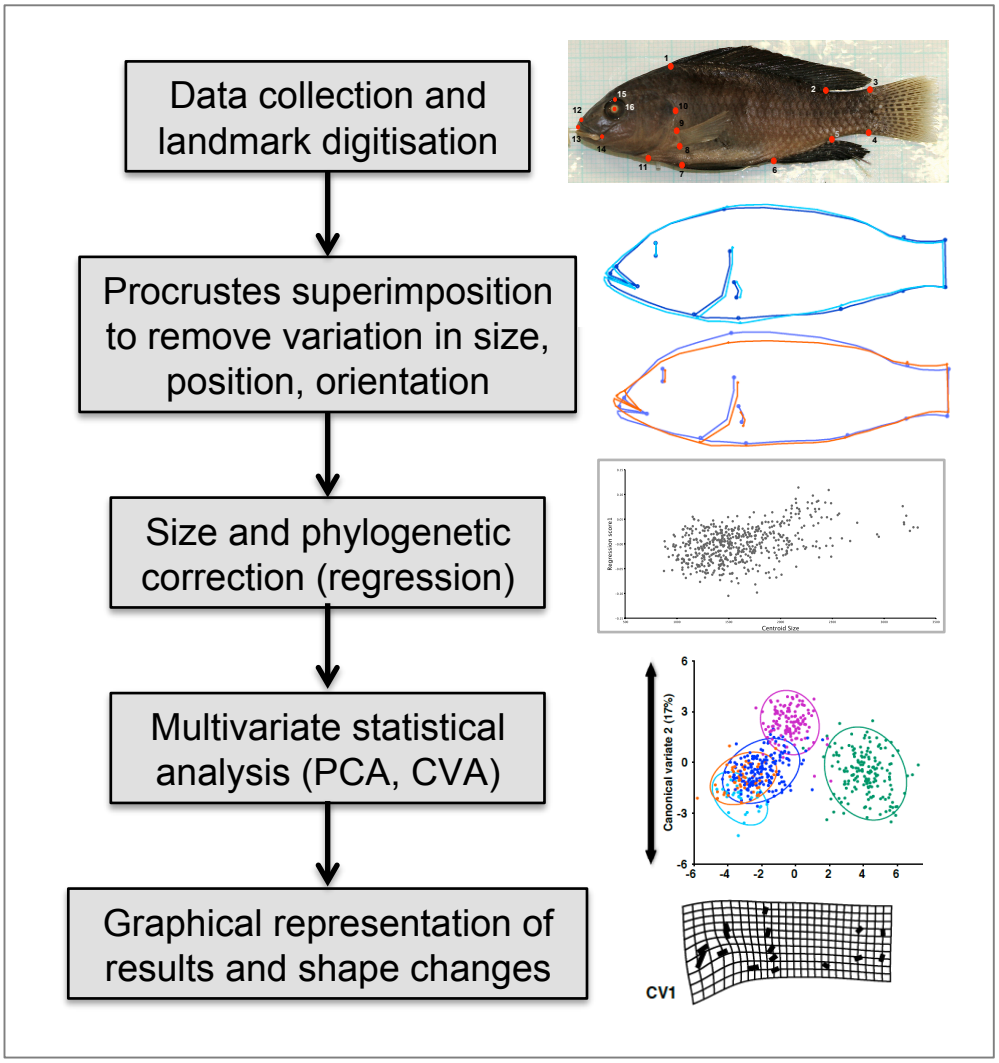


Figure 2.6. Typical workflow for GMM analysis.

In order to compare multiple individuals (represented by multiple landmark configurations) and ensure only variations in shape are analysed, a Procrustes superimposition (Sneath 1967; Rohlf & Slice 1990; Dryden & Mardia 1998) is used to superimpose the configurations and remove variation in size, position, and orientation using a least-squares methodology (reviewed in Klingenberg 2010; 2013). Although such an imposition minimises the absolute differences in size between landmark figurations, it does not account for the correlation between overall size of an organism and its shape. Body shape often has a strong allometric component (e.g., Sidlauskas *et al.* 2011), so analysis will typically focus on adult individuals to avoid ontogenetic effects (Klingenberg 2013). Conversely, studies investigating shape variation in development or ontogeny will include the widest range in individual size to maximise allometric effect. Allometric effects are more likely where multiple groups are compared, as size composition of each sample may vary, and particularly across different species as allometry varies with ontogeny and

phylogeny (Gould 1966). As such, where allometry would be a confounding factor a correction may be performed whereby the component of shape that is due to variation in size is removed using a regression of shape (e.g., Procrustes coordinates) on size (which may be a separate measurement such as body length, but is more typically the centroid size of the landmark configuration, as method of size correction can have a large effect; Adams *et al.* 2002), and the resulting residuals used in downstream analyses (Monteiro 1999; Klingenberg 2010). Finally, the effects of shared evolutionary history may also need to be removed before analysis. Geometric morphometric analyses often consider variation within species, but for inter-species studies, phylogenetic information needs to be considered alongside morphometric data (Klingenberg & Marugan-Lobon 2013) as closely related species would be expected to be more similar in body shape. Thus, similarities in shape due to ancestry must be excluded before other variables such as environmental conditions are considered (Felsenstein 1985). Data may be corrected for phylogeny using several methods including the use of residuals results from phylogenetic generalised least squares models, phylogenetic independent contrasts, multivariate regression or partial least squares (reviewed in Monteiro 2013). More recent approaches allow for simultaneous correction of size and phylogeny and methods for estimating principal components of shape data that account for phylogenetic non-independence (phylogenetic PCA) (Revell 2009; 2011; Monteiro 2013; Polly *et al.* 2013). However, it should be noted that while pPCA conducts PCA taking phylogeny into account, it does not produce phylogenetically corrected PC scores (D Polly, pers. comm.).

Several statistical techniques have been employed to explore the variation of morphometric data, including PCA, discriminant function analysis (DFA) and the multiple group equivalent canonical variate analysis (CVA). The latter two analyses have caused controversy in terms of validity for morphometric data (Klingenberg & Monteiro 2005), with some authors arguing that the data should only be analysed using absolute values such as Procrustes coordinates (as used in PCA) (Bookstein 1991). However, measures of variation observed relative to the total variation (i.e., DFA and CVA) are frequently used by researchers comparing groups (e.g., Rüber & Adams 2001; Klingenberg *et al.* 2003; Francuski *et al.* 2009a; Arnegard *et al.* 2010; Odhiambo *et al.* 2011). Discriminant function analysis and CVA specifically address multiple group comparisons, while PCA is inherently a single group procedure and thus may fail to find group differences as it does not take group structure into consideration (Strauss 2010). While PCA redistributes data points on orthogonal axes such that variances of scores of individuals are maximised on each axis for the

whole dataset, DFA and CVA maximise discrimination of groups by ensuring maximum between-group variance on each axis (and minimising within-group variation). These approaches are thus preferable when considering differences between groups and attempting to find axes of variation between those groups, but require *a priori* group assignation and can result in artificial discrimination of groups when sample sizes are small (Strauss 2010). Thus, both PCA and CVA may be used in tandem dependent on the questions being addressed (Harvati 2003; Maderbacher *et al.* 2008; Spreitzer *et al.* 2011; Wanek & Sturmbauer 2015; Su *et al.* 2015). A related question arises in the comparison on group means, as to whether group variance is accounted for. Morphological distance is often represented as the distance between group means, but using the Euclidean (or Procrustes) distance can obscure patterns in the data if there are large relative amounts of within-group variation. As such, data is often transformed to standardise within-group variance and the transformed distances are used instead, such as the Mahalanobis distance (Figure 2.7; Klingenberg & Monteiro 2005).

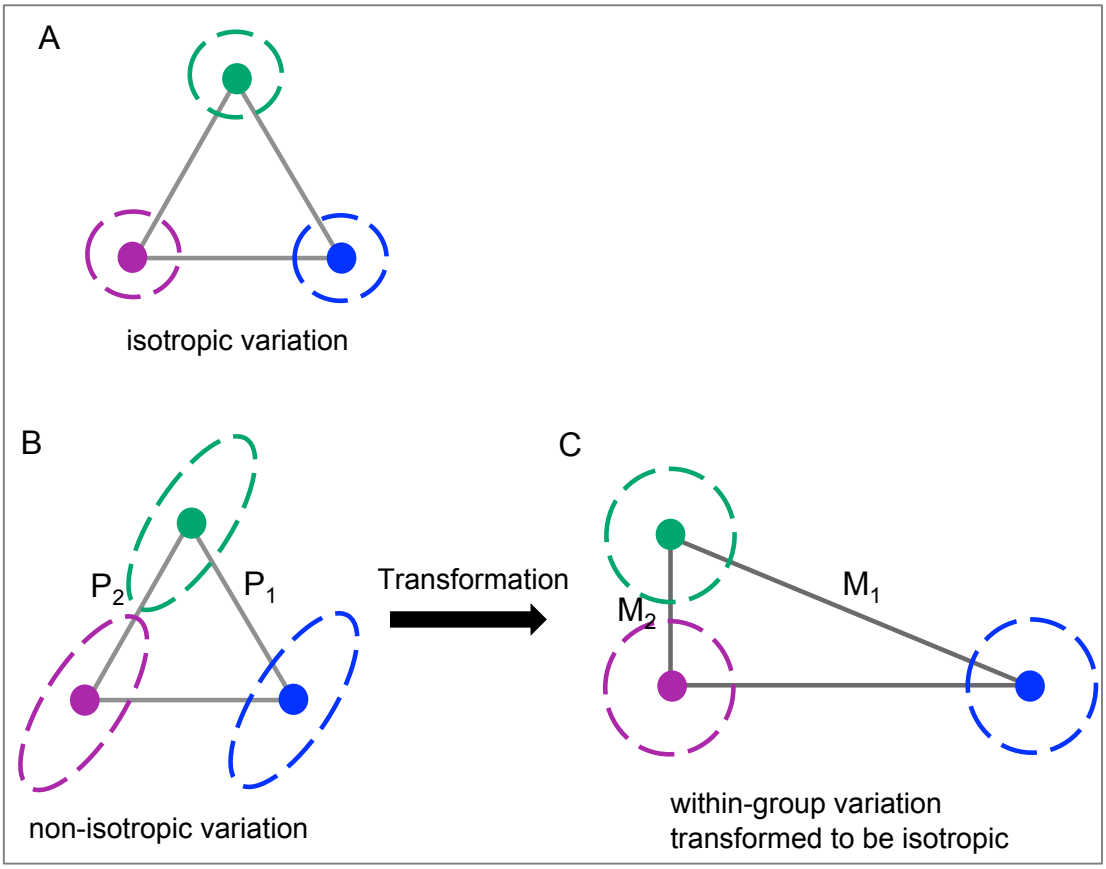


Figure 2.7. Description of group distance measures.

Adapted from Klingenberg & Monteiro (2005). A) Three groups with isotropic variation. Variation around the mean (dotted line) is equal around each group mean (solid circle) and Procrustes (Euclidean) distances between each group mean (grey lines) are equal. B) Three groups with non-isotropic variation. Although the means are all equally far apart measured by Procrustes distance (P1 and P2), variation around the mean suggests that the green and pink groups are less distinct from each other than from the blue group – an observation that would be obscured if only considering the group means and Procrustes distance. C) Panel in (B) transformed to make variation around each mean isotropic, with Procrustes distances transformed to Mahalanobis distances, such that the distance between the pink and green groups (M1) is no longer the same as that between green and blue groups (M2).

The development of geometric morphometrics and correlation of landmarks represented a significant advance in the analysis of organismal form (Corti 1993; Rohlf & Marcus 1993; Adams *et al.* 2002), and provided several advantages over traditional morphometrics of independent measurements, especially in the area of visualising shape change (Klingenberg 2013). However, a caveat for the use of landmarks is that shifts of superimposed landmarks cannot be interpreted on a standalone basis (Klingenberg 2013), as the landmark configuration has been rotated to minimise variation at all landmarks during the superimposition process. Using a least-squares method (e.g., General Procrustes) for superimposition is

sensitive to outliers, as a single landmark with large variation would result in the variation being distributed across all landmarks and all being shifted following superimposition. This issue is termed the Pinocchio effect, and where instances of large localised differences are suspected can be addressed by using alternative superimposition methods such as resistant-fit criterion using medians (Zelditch *et al.* 2012).

The current description and species assignation of *Alcolapia* flock is largely dependent on morphology (Trewavas 1983; Seegers & Tichy 1999). Fish body shape not only provides information on evolutionary relationships – being constrained by a strong genetic component (Leionen *et al.* 2011) – but is also influenced by environmental conditions, in particular salinity, water chemistry and flow (Gomes & Monteiro 2008; Firmat *et al.* 2012). Thus, analysis and comparison of the external body shape of the focal fish here may elucidate the relationships between species of *Alcolapia* flock, as well as the phenotype-environment correlation between species.

Study sample sizes

Generally, throughout the present study, species from different populations were grouped for overall species analyses, when population differences were not being investigated or taken into account. However, to avoid geographic or associated biases, where appropriate analysis was also conducted on a smaller subset of data from single populations or only for species occurring in sympatric populations to obviate any environmental differences between sites. Furthermore, while the initial sampling was conducted for phylogenetic purposes, the number of samples collected allowed investigation of other partitioning of data besides species-levels differentiation. However, the sampling was not designed to consider morph or sex differences within species, so it is acknowledged that the unequal sample sizes of these groups may weaken statistical analysis for these additional comparisons.

Integrated analysis

The methods thus far discussed not only address questions of phylogenetics, ecology and development, but also are increasingly being combined to consider evolutionary relationships and model ecological traits over time to infer rates of change, ancestral reconstruction, convergent evolution of body shape, and investigate models of adaptive radiation. Tests for covariation in datasets e.g.,

Mantel tests, partial Mantel tests; regression; and correlation of phenotype with environment, allow for the variation in ecology and morphology with genetics to be tested, and for external environmental variables to infer the axes on which selection pressure may be operating. The original developers of geometric morphometric methods foresaw the importance of this integration of approaches, stating "*It is hoped that, in future, geometric morphometrics will be combined with genetic and ecological knowledge of organisms – a new synthesis in biology*." (Corti 1993). This has certainly been the case in recent years, with several large-scale studies of adaptive radiations employing an interdisciplinary approach of geometric morphometric, ecological (typically diet) and genomic data to address questions such as convergence (Colombo *et al.* 2012; Muschick *et al.* 2012), parallel evolution (Manousaki *et al.* 2013; Elmer *et al.* 2014), stalled speciation (Martin 2013), genetic architecture of niche differentiation (Arnegard *et al.* 2014), ecological speciation (Praebel *et al.* 2013; Martin & Feinstein 2014), and models of adaptive radiation (Wilson *et al.* 2013). It is increasingly recognised that while genes and DNA sequences often reflect complex phylogenetic histories, it is whole organisms that are the targets of selection, leading to a holistic view considering the biological relevance of the patterns observed in genomic data.

References

Adams DC, Rohlf FJ, Slice DE (2002) Geometric morphometrics: Ten years of progress following the “revolution.” *Italian Journal of Zoology*, **71**, 5–16.

Anseeuw D, Guelinckx J, Snoeks J (2009) Differences in stable isotope composition within and among zooplanktivorous Utaka cichlid populations from Lake Malawi. *Aquatic Ecology*, **48**, 378–385.

Arnegard ME, McGee MD, Matthews B *et al.* (2014) Genetics of ecological divergence during speciation. *Nature*, **511**, 307–311.

Arnegard ME, McIntyre PB, Harmon LJ *et al.* (2010) Sexual signal evolution outpaces ecological divergence during electric fish species radiation. *The American Naturalist*, **176**, 335–356.

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Bocher P, Cherel Y, Hobson K (2000) Complete trophic segregation between South Georgian and common diving petrels during breeding at Iles Kerguelen. *Marine Ecology Progress Series*, **208**, 249–264.

Boecklen WJ, Yarnes CT, Cook BA, James AC (2011) On the use of stable isotopes in trophic ecology. *Annual Review of Ecology, Evolution, and Systematics*, **42**, 411–440.

Bookstein FL (1984) Tensor biometrics for changes in cranial shape. *Annals of Human Biology*, **11**, 413–437.

Bookstein FL (1989) Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2**, 567–585.

Bookstein FL (1991) *Morphometric Tools for Landmark Data*. Cambridge University Press, New York.

Bootsma HA, Hecky RE, Hesslein RH, Turner GF (1996) Food partitioning among Lake Malawi nearshore fishes as revealed by stable isotope analyses. *Ecology*, **77**, 1286–1290.

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, **32**, 314–331.

Brito PH, Edwards SV (2008) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica*, **135**, 439–455.

Caut S, Angulo E, Courchamp F (2009) Variation in discrimination factors (Δ^{15}N and Δ^{13}C): The effect of diet isotopic values and applications for diet reconstruction. *Journal of Applied Ecology*, **46**, 443–453.

Chevasco V, Elzinga JA, Mappes J, Grapputo A (2014) Evaluation of criteria for species delimitation of bagworm moths (Lepidoptera: Psychidae). *European Journal of Entomology*, **111**, 121–136.

Coe MJ (1966) The biology of *Tilapia grahami* Boulenger in Lake Magadi, Kenya. *Acta Tropica*, **23**, 146–198.

Colombo M, Diepeveen ET, Muschick M *et al.* (2012) The ecological and genetic basis of convergent thick-lipped phenotypes in cichlid fishes. *Molecular Ecology*, **22**, 670–684.

Connolly RM, Guest MA, Melville AJ, Oakes JM (2004) Sulfur stable isotopes separate producers in marine food-web analysis. *Oecologia*, **138**, 161–167.

Corti M (1993) Geometric morphometrics: An extension of the revolution. *Trends in Ecology & Evolution*, **8**, 302–303.

DaCosta JM, Sorenson MD (2014) Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. *PLoS ONE*, **9**, e106713.

Davey JW, Cezard T, Fuentes-Utrilla P *et al.* (2012) Special features of RAD sequencing data: implications for genotyping. *Molecular Ecology*, **22**, 3151–3164.

Davey JW, Davey JL, Blaxter ML, Blaxter MW (2010) RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, **9**, 416–423.

Davey JW, Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next generation sequencing. *Nature Reviews Genetics*, **12**, 499–510.

Day-Williams AG, Zeggini E (2010) The effect of next-generation sequencing technology on complex trait research. *European Journal of Clinical Investigation*, **41**, 561–567.

DeNiro MJ, Epstein S (1978) Influence of diet on the distribution of carbon isotopes in animals. *Geochimica et Cosmochimica Acta*, **42**, 495–506.

DeNiro MJ, Epstein S (1981) Influence of diet on the distribution on nitrogen isotopes in animals. *Geochimica et Cosmochimica Acta*, **45**, 241–351.

Dryden IL, Mardia KV (1998) *Statistical shape analysis*. John Wiley & Sons, Ltd, Chichester.

Duponchelle F, Ribbink AJ, Msukwa A *et al.* (2005) Food partitioning within the species-rich benthic fish community of Lake Malawi , East Africa. *Canadian Journal of Fisheries and Aquatic Sciences*, **1664**, 1651–1664.

Elmer KR, Fan S, Kusche H *et al.* (2014) Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel. *Nature Communications*, **5**, 1–8.

Emerson KJ, Merz CR, Catchen JM *et al.* (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences*, **107**, 16196–16200.

Felsenstein J (1985) Phylogenies and the comparative method. *The American Naturalist*, **125**, 1–15.

Firmat C, Schliewen UK, Losseau MEL, Alibert P (2012) Body shape differentiation at global and local geographic scales in the invasive cichlid *Oreochromis mossambicus*. *Biological Journal of the Linnean Society*, **105**, 369–381.

Ford AGP, Dasmahapatra KK, Rüber L *et al.* (2015) High levels of interspecific gene flow in an endemic cichlid fish adaptive radiation from an extreme lake environment. *Molecular Ecology*, doi: 10.1111–mec.13247.

Fox EJ, Reid-Bayliss KS, Emond MJ, Loeb LA (2014) Accuracy of next generation sequencing platforms. *Next Generat Sequences & Applications*, **1**, 106.

France RL (1995) Carbon-13 enrichment in benthic compared to planktonic algae: foodweb implications. *Marine Ecology Progress Series*, **124**, 307–312.

Francuski L, Ludoški J, Vujić A, Milankov V (2009a) Wing geometric morphometric inferences on species delimitation and intraspecific divergent units in the *Merodon ruficornis* group (Diptera, Syrphidae) from the Balkan Peninsula. *Zoological Science*, **26**, 301–308.

Francuski L, Vujić A, Kovačević A, *et al*. (2009b) Identification of the species of the *Cheilosia variabilis* group (Diptera, Syrphidae) from the Balkan Peninsula using wing geometric morphometrics, with the revision of status of *C. melanopa redi* Vujić, 1996. *Contributions to Zoology*, **78**, 129–140.

Genner MJ, Turner GF (2012) Ancient hybridization and phenotypic novelty within Lake Malawi's cichlid fish radiation. *Molecular Biology and Evolution*, **29**, 195–206.

Genner MJ, Turner GF, Barker S, Hawkins SJ (1999) Niche segregation among Lake Malawi cichlid fishes? Evidence from stable isotope signatures. *Ecology Letters*, **2**, 185–190.

Gomes JJ, Monteiro LR (2008) Morphological divergence patterns among populations of *Poecilia vivipara* (Teleostei Poeciliidae): Test of an ecomorphological paradigm. *Biological Journal of the Linnean Society*, **93**, 799–812.

Gould SJ (1966) Allometry and size in ontogeny and phylogeny. *Biological Reviews*, **41**, 587–640.

Graham CF, Glenn TC, McArthur AG *et al.* (2015) Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources*, doi: 10.1111–1755–0998.12404.

Grey J, Deines P (2005) Differential assimilation of methanotrophic and chemoautotrophic bacteria by lake chironomid larvae. *Aquatic Microbial Ecology*, **40**, 61–66.

Grey J, Kelly A, Jones RI (2004) High intraspecific variability in carbon and nitrogen stable isotope ratios of lake chironomid larvae. *Limnology and Oceanography*, **49**, 239–244.

Gurgel-Gonçalves R, Ferreira JBC, Rosa AF *et al*. (2010) Geometric morphometrics and ecological niche modelling for delimitation of near-sibling triatomine species. *Medical and Veterinary Entomology*, **25**, 84–93.

Harvati K (2003) Quantitative analysis of Neanderthal temporal bone morphology using three-dimensional geometric morphometrics. *American Journal of Physical Anthropology*, **120**, 323–338.

Henning F, Lee HJ, Franchini P, Meyer A (2014) Genetic mapping of horizontal stripes in Lake Victoria cichlid fishes: benefits and pitfalls of using dense linkage mapping in non-model organisms. *Molecular Ecology*, **23**, 5224–5240.

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.

Huang H, Knowles LL (2014) Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. *Systematic Biology*, **0**, 1–9.

Keeley JE, Sandquist DR (1992) Carbon: freshwater plants. *Plant, Cell and Environment*, **15**, 1021–1035.

Kendall DG (1984) Shape manifolds, Procrusteam metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, **16**, 81–121.

Klingenberg CP (2010) Evolution and development of shape: integrating quantitative approaches. *Nature Reviews Genetics*, **11**, 623–635.

Klingenberg CP (2013) Visualizations in geometric morphometrics: how to read and how to make graphs showing shape changes. *Hystrix*, **24**, 15–24.

Klingenberg CP, Marugan-Lobon J (2013) Evolutionary covariation in geometric morphometric data: Analyzing integration, modularity and allometry in a phylogenetic context. *Systematic Biology*, **62**, 591–610.

Klingenberg CP, Monteiro L (2005) Distances and directions in multidimensional shape spaces: implications for morphometric applications. *Systematic Biology*, **54**, 678–688.

Klingenberg CP, Barluenga M, Meyer A (2003) Body shape variation in cichlid fishes of the *Amphilophus citrinellus* species complex. *Biological Journal of the Linnean Society*, **80**, 397–408.

Layman CA, Arrington DA, Montan CG, Post DG (2007) Can stable isotope ratios provide for community-wide measures of trophic structure? *Ecology*, **88**, 42–48.

Leionen T, Cano JM, Merila J (2011) Genetics of body shape and armour variation in threespine sticklebacks. *Journal of Evolutionary Biology*, **24**, 206–218.

Lemmon EM, Lemmon AR (2013) High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics*, **44**, 99–121.

Lewis ZA, Shiver AL, Stiffler N *et al.* (2007) High-density detection of restriction-site-associated DNA markers for rapid mapping of mutated loci in *Neurospora*. *Genetics*, **177**, 1163–1171.

Liu L, Li Y, Li S *et al.* (2012) Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*, **2012**, 1–11.

Maderbacher M, Bauer C, Herler J *et al.* (2008) Assessment of traditional versus geometric morphometrics for discriminating populations of the *Tropheus moorii* species complex (Teleostei: Cichlidae), a Lake Tanganyika model for allopatric speciation. *Journal of Zoological Systematics and Evolutionary Research*, **46**, 153–161.

Manousaki T, Hull PM, Kusche H *et al.* (2013) Parsing parallel evolution: ecological divergence and differential gene expression in the adaptive radiations of thick-lipped Midas cichlid fishes from Nicaragua. *Molecular Ecology*, **22**, 650–669.

Marijnissen SAE, Michel E, Cleary DFR, McIntyre PB (2008) Ecology and conservation status of endemic freshwater crabs in Lake Tanganyika, Africa. *Biodiversity and Conservation*, **18**, 1555–1573.

Martin CH (2013) Strong assortative mating by diet, color, size, and morphology but limited progress toward sympatric speciation in a classic example: Cameroon crater lake cichlids. *Evolution*, **67**, 2114–2123.

Martin CH, Feinstein LC (2014) Novel trophic niches drive variable progress towards ecological speciation within an adaptive radiation of pupfishes. *Molecular Ecology*, **23**, 1846–1862.

Martin CH, Cutler JS, Friel JP *et al.* (2015) Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. *Evolution*, doi: 10.1111–evo.12674.

Martin SH, Dasmahapatra KK, Nadeau NJ *et al.* (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, **23**, 1817–1828.

McKinney CR, McCrea JM, Epstein S, Allen HA, Urey HC (1950) Improvements in mass spectrometers for the measurement of small differences in isotope abundance ratios. *Review of Scientific Instruments*, **21**, 724.

Middelburg JJ (2014) Stable isotopes dissect aquatic food webs from the top to the bottom. *Biogeosciences*, **11**, 2357–2371.

Miller MR, Atwood TS, Eames BF *et al.* (2007a) RAD marker microarrays enable rapid mapping of zebrafish mutations. *Genome Biology*, **8**, R105.

Miller MR, Dunham JP, Amores A *et al*. (2007b) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, **17**, 240–248.

Monteiro L (1999) Multivariate regression models and geometric morphometrics: The search for causal factors in the analysis of shape. *Systematic Biology*, **48**, 192–199.

Monteiro L (2013) Morphometrics and the comparative method: studying the evolution of biological shape. *Hystrix*, **24**, 25–32.

Muschick M, Indermaur A, Salzburger W (2012) Convergent evolution within an adaptive radiation of cichlid fishes. *Current Biology*, **22**, 1–7.

Mutanen M, Pretorius E (2007) Subjective visual evaluation vs. traditional and geometric morphometrics in species delimitation: a comparison of moth genitalia. *Systematic Entomology*, **32**, 371–386.

Neill C, Cornwell JC (1992) Stable carbon, nitrogen, and sulfur isotopes in a prairie marsh food web. *Wetlands*, **12**, 217–224.

Newsome SD, Martinez C, Bearhop S, Phillips DL (2007) A niche for isotopic ecology. *Frontiers in Ecology and the Environment*, **5**, 429–436.

Newton J (2010) *Stable isotope ecology*. John Wiley & Sons, Ltd, Chichester.

Odhiambo EA, Kerschbaumer M, Postl L, Sturmbauer C (2011) Morphometric differentiation among haplochromine cichlid fish species of a satellite lake of Lake Victoria. *Journal of Zoological Systematics and Evolutionary Research*, **49**, 216–223.

Pace ML, Cole JJ, Carpenter SR *et al.* (2004) Whole-lake carbon-13 additions reveal terrestrial support of aquatic food webs. *Nature*, **427**, 240–243.

Peterson BK, Weber JN, Kay EH *et al*. (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, **7**, e37135.

Polly PD, Lawing AM, Fabre A-C, Goswami A (2013) Phylogenetic principal components analysis and geometric morphometrics. *Hystrix*, **24**, 33–41.

Post DM (2002) Using stable isotopes to estimate trophic position: Models, methods, and assumptions. *Ecology*, **83**, 703–718.

Praebel K, Knudsen R, Siwertsson A *et al.* (2013) Ecological speciation in postglacial European whitefish: Rapid adaptive radiations into the littoral, pelagic, and profundal lake habitats. *Ecology and Evolution*, **3**, 4970–4986.

Revell LJ (2009) Size-correction and principal components for interspecific comparative studies. *Evolution*, **63**, 3258–3268.

Revell LJ (2011) phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**, 217–223.

Rohlf FJ, Marcus LF (1993) A revolution in morphometrics. *Trends in Ecology & Evolution*, **8**, 129–132.

Rohlf FJ, Slice D (1990) Extensions of the procrustes method for the optimal superimposition of landmarks. *Systematic Zoology*, **39**, 40–59.

Rubin BER, Ree RH, Moreau CS (2012) Inferring phylogenies from RAD sequence data. *PLoS ONE*, **7**, e33394.

Rüber L, Adams DC (2001) Evolutionary convergence of body shape and trophic morphology in cichlids from Lake Tanganyika. *Journal of Evolutionary Biology*, **14**, 325–332.

Schwarzfeld M, Sperling F (2014) Species delimitation using morphology, morphometrics, and molecules: definition of the *Ophion scutellaris* Thomson species group, with descriptions of six new species (Hymenoptera, Ichneumonidae). *ZooKeys*, **462**, 59–114.

Seegers L, Sonnenberg R, Yamamoto R (1999) Molecular analysis of the *Alcolapia* flock from lakes Natron and Magadi, Tanzania and Kenya (Teleostei: Cichlidae), and implications for their systematics and evolution. *Ichthyological Explorations of Freshwaters*, **10**, 175–199.

Shaffer HB, Thomson RC (2007) Delimiting species in recent radiations. *Systematic Biology*, **56**, 896–906.

Sidlauskas BL, Mol JH, Vari RP (2011) Dealing with allometry in linear and geometric morphometrics: a taxonomic case study in the *Leporinus cylindriformis* group (Characiformes: Anostomidae) with description of a new species from Suriname. *Zoological Journal of the Linnean Society*, **162**, 103–130.

Smyntek PM, Maberly SC, Grey J (2012) Dissolved carbon dioxide concentration controls baseline stable carbon isotope signatures of a lake food web. *Limnology and Oceanography*, **57**, 1292–1302.

Sneath PHA (1967) Trend-surface analysis of transformatrion grids. *Journal of Zoology*, **151**, 65–122.

Spreitzer ML, Mautner S, Makasa L, Sturmbauer C (2011) Genetic and morphological population differentiation in the rock-dwelling and specialized shrimp-feeding cichlid fish species *Altolamprologus compressiceps* from Lake Tanganyika, East Africa. *Hydrobiologia*, **682**, 143–154.

Stapley J, Reger J, Feulner PGD *et al.* (2010) Adaptation genomics: the next generation. *Trends in Ecology & Evolution*, **25**, 705–712.

Strauss RE (2010) Discriminating groups of organisms. In: *Lecture Notes in Earth Sciences*. pp. 73–91. Springer Berlin Heidelberg, Berlin.

Su J, Guan K, Wang J, Yang Y (2015) Significance of hind wing morphology in distinguishing genera and species of cantharid beetles with a geometric morphometric analysis. *ZooKeys*, **502**, 11–25.

Tichy H, Seegers L (1999) The *Oreochromis alcalicus* flock (Teleostei Cichlidae) from lakes Natron and Magadi, Tanzania and Kenya: A model for the evolution of new species flocks in historical times. *Ichthyological Explorations of Freshwaters*, **10**, 147–174.

Toonen RJ, Puritz JB, Forsman ZH *et al.* (2013) ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ*, **1**, e203.

Vander Zanden MJ, Clayton MK, Moody EK *et al*. (2015) Stable isotope turnover and half-life in animal tissues: A literature synthesis. *PLoS ONE*, **10**, e0116182.

Vander Zanden MJ, Shuter BJ, Lester N, Rasmussen JB (1999) Patterns of food chain length in lakes: A stable isotope study. *The American Naturalist*, **154**, 406–416.

Vanderklift MA, Ponsard S (2003) Sources of variation in consumer-diet $\delta^{15}N$ enrichment: a meta-analysis. *Oecologia*, **136**, 169–182.

Vos P, Hogers R, Bleeker M *et al.* (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, **23**, 4407–4414.

Wagner CE, Keller I, Wittwer S *et al.* (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**, 787–798.

Wagner CE, McCune AR, Lovette IJ (2012) Recent speciation between sympatric Tanganyikan cichlid colour morphs. *Molecular Ecology*, **21**, 3283–3292.

Wanek KA, Sturmbauer C (2015) Form, function and phylogeny: comparative morphometrics of Lake Tanganyika's cichlid tribe Tropheini. *Zoologica Scripta*, doi:10.1111–zsc.12110.

Wang S, Meyer E, McKay JK, Matz MV (2012) 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods*, **9**, 808–810.

Weidel BC, Carpenter SR, Kitchell JF, Vander Zanden MJ (2011) Rates and components of carbon turnover in fish muscle: insights from bioenergetics models and a whole-lake ^{13}C addition. *Canadian Journal of Fisheries and Aquatic Sciences*, **68**, 387–399.

Wilson LAB, Colombo M, Hanel R, *et al*. (2013) Ecomorphological disparity in an adaptive radiation: opercular bone shape and stable isotopes in Antarctic icefishes. *Ecology and Evolution*, **3**, 3166–3182.

Wilson PJ, Wood CM, Walsh PJ *et al.* (2004) Discordance between genetic structure and morphological, ecological, and physiological adaptation in Lake Magadi tilapia. *Physiological and Biochemical Zoology*, **77**, 537–555.

Won Y-J, Wang Y, Sivasundar A, *et al*. (2006) Nuclear gene variation and molecular dating of the cichlid species flock of Lake Malawi. *Molecular Biology and Evolution*, **23**, 828–837.

Zelditch ML, Swiderski D, Sheets HD (2012) *Geometric morphometrics for biologists: A primer*. Academic Press, London.

Appendix 2.1 – Preliminary genetic analysis, methods

Preliminary genetic analysis included specimens of *Alcolapia* and related *Oreochromis* specimens for analysis of the mitochondrial control region (CR). The CR was amplified and sequenced for 43 individuals: *A. ndalalani* (n=6), *A. alcalica* (n=16), *A. latilabris* (n=10), *O. amphimelas* (n=4), *Oreochromis spp.* (n=6), and *Haplochromis spp.* (n=1).

Extraction and amplification conditions

Genomic DNA was extracted from fin clip samples stored in 95% ethanol using the Qiagen DNeasy Blood and Tissue kit. The CR was amplified using primers L-PRO-F (Meyer *et al.* 1994) and TDK-DHG (Lee *et al.* 1995) in the following reaction mixture: 2 µL buffer, 0.8 µL $MgCl_2$, 0.8 µL dNTPs, 1.0 µL of each primer, 0.08 µL Taq polymerase, 2.0 µL of DNA sample, 12.32 µL H_2O to a total reaction volume of 20 µL. The following PCR cycle was used: 1 minute at 94˚C, followed by 34 cycles of (35 seconds at 94˚C, 35 seconds at 54˚C, 90 seconds at 72˚C), followed by a final 4-minute 72˚C extension step. PCR products were checked for successful amplification by electrophoresis on a 1.5% agarose gel using GelRed stain and visualised under UV light. PCR products were cleaned using MicroClean (5mL 5M NaCl, 0.1mL 1M Tris-HCl, 0.02mL 0.5M EDTA, 20g PEG 8000, 0.086mL 2M $MgCl_2$, 50 mL distilled H_2O), and were sequenced on an ABI 3730 sequencer in forward and reverse directions using the original PCR primers and two internal primers: TDK-D (Lee *et al.* 1995) and SC-DL (Salzburger *et al*. 2002).

Molecular analysis

Sequences were checked by eye, primer sequences removed and contigs joined in Sequencher 4.8 (Gene Codes Corporation). Sequences were aligned in MEGA 5.05 (Tamura *et al.* 2011) using MUSCLE (Edgar 2004), which is slightly more accurate than ClustalW alignment (Nuin *et al.* 2006). Default parameters for the MUSCLE alignment were used (Gap Open penalty: -400; Gap extend penalty: 0; clustering method: UPGMA), and pairwise distances were calculated in MEGA 5.05 using default parameters to check suitable quality, i.e., <0.33 pairwise difference between sequences for non-coding regions (Kumar & Filipski 2007; Hall 2011). Maximum likelihood inference was implemented using modified parameters in MEGA. The MEGA test of likelihood selected TN93+G+I as the best evolutionary model, and the

analysis was rooted using *Haplochromis* spp. Bayesian analysis was conducted in Mr Bayes 3.2 (Ronquist *et al.* 2012) using default parameters with an invariant gamma rate distribution and run for $10x10^6$ generations sampling every 100 generations (four chains, temperature 0.2), with the first 250,000 generations discarded as burn in.

Appendix references

Edgar RC (2004) MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, **19**, 1–19.

Hall BG (2011) *Phylogenetic Trees Made Easy: A How-To Manual*. Sinauer Associates, Sunderland, MA.

Kumar S, Filipski A (2007) Multiple sequence alignment: in pursuit of homologous DNA positions. *Genome Research*, **17**, 127–135.

Lee WJ, Conroy J, Howell WH, Kocher TD (1995) Structure and evolution of teleost mitochondrial control regions. *Journal of Molecular Evolution*, **41**, 54–66.

Meyer A, Morrissey JM, Schartl M (1994) Recurrent origin of a sexually selected trait in *Xiphophorus* fishes inferred from a molecular phylogeny. *Nature*, **368**, 539–542.

Nuin PAS, Wang Z, Tillier ERM (2006) The accuracy of several multiple sequence alignment programs for proteins. *BMC Bioinformatics*, **7**, 471.

Ronquist F, Teslenko M, van der Mark P *et al.* (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539-542.

Salzburger W, Meyer A, Baric S, Verheyen E, Sturmbauer C. (2002) Phylogeny of the Lake Tanganyika cichlid species flock and its relationship to the Central and East African haplochromine cichlid fish faunas. Systematic Biology, **51**, 113–135.

Tamura K, Peterson D, Peterson N *et al.* (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Chapter three

Phylogeography of the *Alcolapia* species flock

Abstract

Studying recent adaptive radiations in isolated insular systems avoids complicating causal events and thus may offer clearer insight into mechanisms generating biological diversity. Here evolutionary relationships and genomic differentiation within the recent radiation of the unique *Alcolapia* cichlid fishes are investigated. Analysis of an extensive RAD dataset of 91 individuals across multiple sampling sites finds evidence for genetic admixture between species within Lake Natron, with the highest levels of admixture between sympatric populations of the most recently diverged species. Despite considerable environmental separation, populations within Lake Natron do not exhibit isolation by distance, indicating panmixia within the lake, although individuals within lineages cluster by population in phylogenomic analysis. These results indicate exceptionally low genetic differentiation across the radiation despite considerable phenotypic trophic variation, supporting previous findings from smaller datasets. Evidence of ongoing gene flow and interspecies hybridisation in certain populations suggests that *Alcolapia* species are incompletely reproductively isolated.

Introduction

The study of adaptive radiation, the evolution of ecological diversity in rapidly multiplying lineages (Schluter 2000), is important in understanding the diversification of incipient species, as well as the genetic and ecological structure of species diversity (Hudson *et al.* 2010). Adaptive radiations from isolated insular systems, as opposed to more complex systems, offer clearer insight into fundamental evolutionary questions regarding the mechanisms generating biological diversity, and the role of ecological opportunity and sexual selection in the origin of species (e.g., Seehausen 2006; 2013; Gillespie 2013).

A major obstacle to studying recent radiations is insufficient genetic differentiation to define and characterise species relationships. However, the advent of high-throughput sequencing has greatly facilitated the study of shallow

divergence. Several recent studies have demonstrated the successful application of the reduced-representation methodology of RAD (restriction-site associated DNA) sequencing (Baird *et al.* 2008) to phylogenomic reconstruction (Rubin *et al.* 2012; Cariou *et al.* 2013; Eaton & Ree 2013; Wagner *et al.* 2013). Of these, Wagner *et al*. (2013) achieved exceptional phylogenetic resolution for the relatively young Lake Victoria cichlid fish adaptive radiation. These approaches, made possible by advances in sequencing technology, provide greater clarity on the basis for morphological species designation in cases where morphology does not match the molecular phylogeny (Keller *et al.* 2013). Furthermore, species delimitation is difficult in recent radiations using genealogical approaches of single-gene or multi-gene alignments, and analyses indicate that population genomic approaches based on large sets of SNPs are more reliable in delimiting recently derived species (Shaffer & Thomson 2007).

In the present analysis, species and population relationships within the recent, small-scale radiation of *Alcolapia* cichlids are investigated using phylogenomic methods. The endemicity of the *Alcolapia* species flock, together with the fact that they are only fish present in the in the extreme environment in which they occur may make inference of colonisation patterns clearer than in cases where multiple lineages coexist. Given the extremely young age of the radiation (divergence as recently as ~10,000 years ago based on geological evidence Williamson *et al.* 1993; Tichy & Seegers 1999), it also seems likely that these fishes have adapted to extreme conditions rapidly. The soda lake system is similar to other recent, small-scale freshwater fish radiations such as the Neotropical crater lake cichlids (Barluenga & Meyer 2004; Barluenga *et al*. 2006; Elmer *et al.* 2010b; 2012), postglacial lake whitefish (Vonlanthen *et al.* 2009; Praebel *et al.* 2013) and three-spined sticklebacks (Reusch *et al.* 2001; Aguirre *et al.* 2008) regarding its young geological age and highly restricted geographic area. These factors potentially make colonisation inference more straightforward than in larger water bodies such as the African Great Lakes with older or less well-defined geological histories and greater species diversity.

Both of the focal lakes are shallow endorheic basins, with Lake Natron having an average lake area of 398 km^2 varying from 81–804 km^2 (Tebbs *et al.* 2013), and Lake Magadi ranging from 75–108 km^2 in dry to wet seasons (Jones *et al.* 1977; Vanden Bossche & Bernacsek 1990). The lakes are subject to substantial climatic effects, with a negative evaporative balance (Burrough & Thomas 2009). A thick layer of crystalline trona (sodium carbonate precipitate) covers most of the lakes' surfaces, forming a solid covering separating lagoons of permanent open water

close to the shore (Kaufman *et al.* 1990). The area is volcanic with alkaline hydrothermal springs containing high levels of salts and precipitates feeding into the lagoons (Williamson *et al.* 1993). It is along these springs in which *Alcolapia* occur, although they also inhabit edges of the lagoons where the springwater meets the lake body (Narahara *et al.* 1996; Seegers & Tichy 1999). The lagoons are intermittently connected during heavy floods in the rainy season, which may allow migration of *Alcolapia* between populations usually restricted to isolated lagoons during the dry season (Seegers & Tichy 1999; Zaccara *et al.* 2014).

Currently, *Alcolapia* includes four described species (shown in Figure 1), three of which are found within Lake Natron: *Alcolapia alcalica* (Hilgendorf 1905), *Alcolapia latilabris* (Seegers & Tichy 1999) and *Alcolapia ndalalani* (Seegers & Tichy 1999), occurring sympatrically at springs bordering the southern lagoon. *Alcolapia alcalica* is the only species with a lake-wide distribution, also occurring in monospecific populations along the western and northern shores of Lake Natron (Figure 3.1). Additionally, there are colour and trophic morphs found within certain Lake Natron *A. alcalica* populations (Seegers & Tichy 1999; Tichy & Seegers 1999; Seegers *et al.* 2001). The fourth species, *Alcolapia grahami* (Boulenger 1912) is restricted to Lake Magadi and satellite lake Little Magadi (Seegers & Tichy 1999), with a translocated population introduced to Lake Nakuru, Kenya, in the 1950s (Vareschi 1979). The four species exhibit extensive differentiation of trophic morphology, on which original descriptions were based (Trewavas 1983; Seegers & Tichy 1999), with mouth morphology including terminal/retrognathous snout (*A. alcalica*), terminal/prognathous snout (*A. grahami*), inferior thick-lipped (*A. latilabris*) and sub-terminal thin-lipped (*A. ndalalani*) (Trewavas 1983; Seegers & Tichy 1999). The species are also differentiated by breeding male colouration (Trewavas 1983; Seegers & Tichy 1999). Although no formal tests of species status and reproductive isolation have been conducted on *Alcolapia*, aquarium observations indicate preferential mating with conspecifics via female choice (Seegers *et al.* 2001). In mixed Lake Natron species tanks male *A. alcalica* reportedly courted female heterospecifics without any successful spawning (Seegers *et al.* 2001), while hybrid *A. latilabris*/*A. ndalalani* were produced, but only when no male *A. ndalalani* were present and female *A. ndalalani* mated with dominant male *A. latilabris* (Seegers *et al.* 2001). Furthermore, the characteristic trophic morphology of the Natron species was maintained in laboratory populations over several generations (maintained up to F_6), and did not differ in response to food type, indicating a genetic component rather than a plastic response to environment (Seegers & Tichy 1999; Seegers *et al.* 2001).

The conditions in the volcanic springs represent one of the most hostile environments to support fish life, including water temperatures of 30–42.8°C, pH ~10.5, fluctuating dissolved oxygen levels of 0.08-6.46mg/L, and high salt concentrations (>20 ppt). Unique physiological adaptations to alkaline/saline and hypoxic conditions (mostly reported for *A. grahami*, with some limited study of *A. alcalica*) include: ureotelism (Randall *et al*. 1989), specialised gill structure for urea transport and excretion (Narahara *et al*. 1996; Walsh *et al*. 2001), high intracellular pH (Wood *et al*. 1994), a trifurcated oesophagus to prevent alkaline water diluting stomach acid (Bergman *et al.* 2003), and facultative air-breathing via the air bladder (Maina 2000; Johannsson *et al*. 2014). Despite uncertainty over the generic status of *Alcolapia* (discussed in chapter one), for the purposes of the present study the constituent described *Alcolapia* species are considered as valid taxonomic species, following the taxonomy of (Eschmeyer 2015), and employing a phenotypic/cohesion species concept (e.g., Templeton 1989; Mallet 1995).

Lakes Natron and Magadi are situated in a basin that formed 1.7 Ma, and contained a single palaeolake Orolonga from ~700 KYA (Eugster 1986) (Figure 1.3 in chapter one). Based on geological evidence the palaeolake exhibited lower salinity conditions than currently found (Eugster 1986). The separate lakes formed from the palaeolake during an arid event ~11 KYA (Williamson *et al.* 1993), and the hypersaline and alkaline conditions of the current lakes arose ~7 KYA (Roberts *et al.* 1993). Furthermore, fossils found in the high lake-level beds surrounding Lake Magadi and ^{14}C-dated (9,120±170 years) are thought to be of an *Oreochromis* species exhibiting considerably larger body size than present day *Alcolapia* (Coe 1966; Trewavas 1983; Tichy & Seegers 1999). It therefore seems likely that *Alcolapia* adaptation to life in the extreme soda environment and subsequent diversification has occurred within a very short timeframe. *Alcolapia* life history and the hostile environment are both conducive to rapid evolution. The short generation time recorded in these maternal mouth-brooding cichlids (brooding period ~2 weeks, mating within seven weeks; (Coe 1966; Trewavas 1983) is suggested to be a result of the increased metabolic rate required to inhabit elevated temperatures (McCormick *et al.* 2013). Such short generation times together with low effective population sizes allow for rapid allele fixation and drift mechanisms. Moreover, it has been suggested that the conditions not only create a strong selection pressure (Wilson *et al*. 2004), but the extreme environment of alkalinity, free radicals and exposure to UV light may promote new adaptations due to elevated mutation rates (Seegers *et al.* 1999; Portner *et al.* 2010).

A genomic approach is well suited to this system as previous studies have been unable to resolve constituent species using mtDNA (control region 350-450bp; cytb 420bp) or microsatellite nuclear markers. These studies found shared haplotypes between the lakes and species (Seegers *et al*. 1999; Wilson *et al*. 2004; Zaccara *et al*. 2014), but also suggested some separation between the lakes, and an analysis of Lake Magadi (mtDNA 1913bp; nuclear microsatellite loci) indicated structure between *A. grahami* populations (Kavembe *et al.* 2013). A summary of previous molecular analysis of the system is provided in Table 3.1

Aims

Here, *Alcolapia* evolutionary relationships are characterised through dense sampling employing a genome-wide SNP approach. As well as phylogenomically testing species hypotheses, the population connectivity of *Alcolapia* within each of the lakes is tested, and the occurrence of interspecific gene flow within the radiation is examined.

Table 3.1. Previous molecular analyses of the *Alcolapia* cichlid radiation.

Study	Markers	Specimens	Sites sampled	Relationship to *Oreochromis*	*Alcolapia* population structuring	*Alcolapia* population demographics
Sültmann H *et al*. 1995	nuclear: 3 RAPD markers (DXTU1, DXTU2, DXTU3)	n=3 (Aa:2; Ag: 1)	2 (1 Natron; 1 Magadi)	Genus, sister to *Oreochromis*	N/A	N/A
Seegers L *et al*. 1999	mtDNA: tRNAGlu and cytb (417bp) tRNAPro and D-Loop (482bp)	n=64 (Aa: 34; Ag: 13; Al: 8; An: 6)	12 (9 Natron; 3 Magadi)	Genus	Some separation of haplotypes between Northern and Southern locations of Lake Natron. Highest variability at type location	N/A
Wilson PJ *et al.* 2000	mtDNA: CR (350bp) nuclear: VNTR	n=60 (Aa: 20; Ag: 40)	3 (1 Natron; 2 Magadi)	N/A	Significant population structuring between the two Magadi populations.	N/A
Nagl S *et al*. 2001	mtDNA tRNAPro and control region (454bp, masked to 420)	n=114 (Aa: 55; Ag: 25; Al: 25; An: 9)	4 (3 Natron; 1 Magadi)	Clade within *Oreochromis*,	N/A	N/A
Wilson PJ *et al*. 2004	mtDNA: control region (350bp)	n=185 (Aa: 27; Ag: 158)	8 (1 Natron; 6 Magadi; 1 Little Magadi)	N/A	Differentiation between Natron, Magadi, Little Magadi, but low levels of structuring between populations within Magadi. No correlation of gene flow with geography	Evidence of bottlenecks in Magadi populations. Evidence of expansion for Little Magadi / Lake Natron
Kavembe G *et al.* 2013	mtDNA: control (889bp) ND2 (1042bp) 10 microsatellite loci	All *A. grahami* mtDNA n=91 microsatellite n=259	5 (4 Magadi; 1 Little Magadi)	Distinct clade within *Oreochromis*, sister to *O. variabilis* (ND2)	mtDNA: structure between Magadi/ Little Magadi; not within Magadi. Microsatellite: Little Magadi and Fish Springs lagoon form separate clusters. No evidence of isolation by distance	No significant results for past or recent population size change; no evidence of bottlenecks or size reduction
Zaccara S *et al*. 2014	mtDNA: control region (350bp) 7 microsatellite loci	mtDNA n=70 (Aa: 41; Ag: 6; Al: 18; An: 5) microsat n=310 (Aa: 181; Ag: 15; Al: 78; An: 36)	9 (8 Natron; 1 Magadi)	N/A	Limited genetic differentiation between populations, high rate of migrants between sites, no apparent population structuring No evidence of isolation by distance (geographic)	Significant evidence of bottleneck and subsequent expansion for *A. alcalica* and *A. latilabris*

Methods

Taxonomic sampling

Samples were collected in 2012 using hand, cast or seine nets (Appendix Table 3A.1). Fish were euthanised using tricaine methanesulfonate (MS222) and preserved as voucher specimens in 80% ethanol, with genetic samples (fin clips) stored in 95% ethanol. Specimens were identified to species level according to the current taxonomic key (Seegers & Tichy 1999). Sampling locations are shown in Figure 3.1. Straight-line geographic distance between sampling locations were calculated using the Vincenty formula (Vincenty 1975) via GPS Visualizer (http://www.gpsvisualizer.com/calculators), and lake-perimeter distances between sites were estimated using the GPS coordinates plotted in ArcGIS version 10 and summing over distances from intermediate sites.

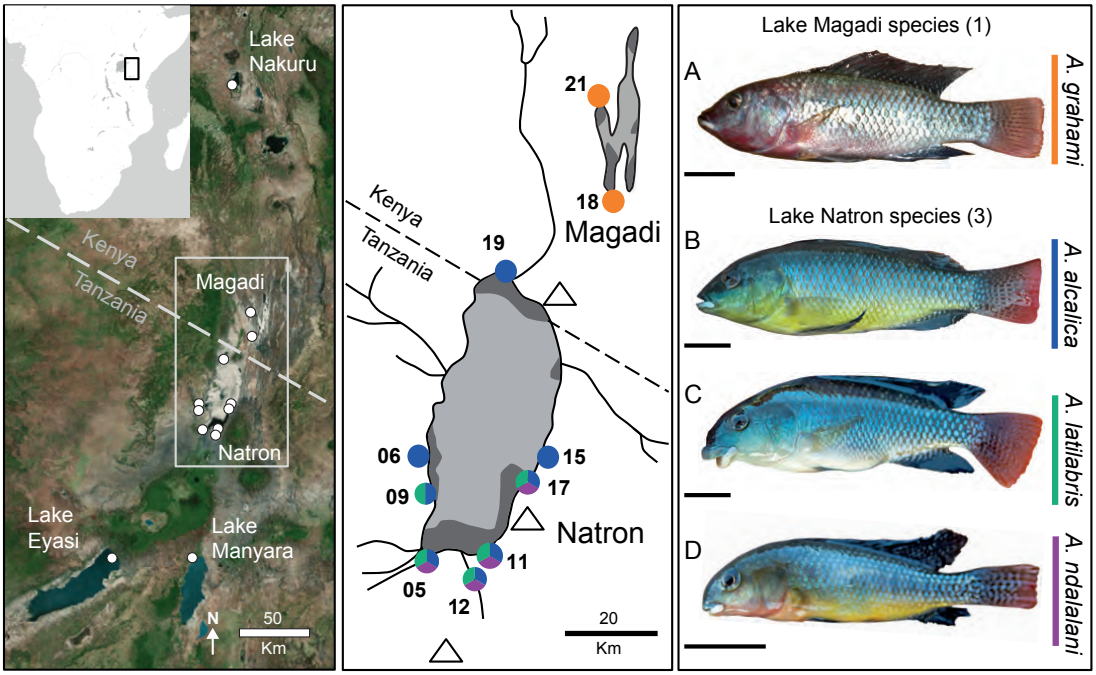


Figure 3.1. Map of soda lake sampling sites. Populations sampled in the present study are shown with white circles in the left-hand panel. In the middle panel, sample site markers are coloured by species present at each site: *A. alcalica*, blue; *A. latilabris*, green; *A. ndalalani*: purple; *A. grahami*: orange. Volcanoes are shown as open triangles; lake extent is shown in light grey, with open water lagoons designated in dark grey. The right-hand panel shows the morphological diversity of the *Alcolapia* described species: A) *A. grahami*; B) *A. alcalica*; C) *A. latilabris*; D) *A. ndalalani*. Scale bar: 10mm. Colour bars to the right indicate colours used for respective species in all figures.

RAD library construction

A total of 96 individuals were sequenced for RAD-tag generation comprising 88 *Alcolapia* specimens, and eight *Oreochromis amphimelas* samples from soda Lakes Manyara and Eyasi (Appendix Table 3A.1) selected as the outgroup (Trewavas 1983; Nagl *et al.* 2001). Genomic DNA was extracted from fin clips, using the DNeasy Blood and Tissue kit (Qiagen). For degraded samples and those for which sufficient yield was not achieved using the Qiagen kits, DNA was extracted using a high-salt chloroform/phenol protocol, (methods in Appendix).

RAD library preparation, sequencing and preliminary bioinformatic processing was undertaken by Edinburgh Genomics (University of Edinburgh). Library preparation followed the protocol of Davey *et al*. (2013), using *SbfI* as the restriction enzyme. Samples were individually barcoded and multiplexed during preparation resulting in a total of six indexed RAD libraries. Libraries were sequenced using a 100 bp paired-end sequencing strategy on Illumina HiSeq 2000 (v3 chemistry).

Libraries were initially sequenced across two lanes, but due to poor initial sequencing quality, each lane was re-sequenced once. One library showed a highly variable number of reads across samples and was therefore prepared again before sequencing in one third of a lane. Reads from all lanes were combined to maximise coverage.

RAD SNP calling

RAD libraries were demultiplexed using the process_radtags function of STACKS v0.99993 (Catchen *et al*. 2011), and individual reads aligned to the *O. niloticus* reference genome v Orenil1.1 (NCBI Assembly GCA_000188235.2, Brawand *et al*. 2014) using the Burrows Wheeler Aligner BWA-backtrack function (Li & Durbin 2009). The resultant sam files were converted to bam files using samtools (Li *et al*. 2009) and duplicate reads marked for removal using picardtools (http://picard.sourceforge.net) to mitigate the effect of biased PCR amplification during library construction. Bam files were realigned around indels using the Genome Analysis Toolkit (GATK) 2.7-2 (McKenna *et al*. 2010). SNP genotyping was carried out using the GATK UnifiedGenotyper (DePristo *et al*. 2011) with default parameters and an emission confidence of 20, and run separately for *O. amphimelas*, *A. alcalica*, *A. grahami*, *A. latilabris* and *A. ndalalani,* with *O. niloticus* specified as the reference genome.

The resultant vcf files were filtered using custom perl scripts as used in (Hoffman *et al.* 2014). Datasets were tested using all variables for the following parameter thresholds: genotype quality: 10, 20 and 30; total bases sequenced per individual (2.5, 5, 10, 15 million bases); missing data threshold: 0%, 25%, 50%, 75%, no threshold. The resultant datasets were analysed using RAxML to assess phylogenetic signal and bootstrap support for trees generated using 100 bootstraps with the GTRGAMMA model. Topology was generally consistent amongst all the datasets (except the most stringent thresholds), however bootstrap support was the highest for the most lenient thresholds (Appendix Figures 3A.1, 3A.2). Thus the following final filtering parameters were used to generate the datasets used in all downstream analyses: SNP quality: 20; genotype quality: 20; mapping quality: 20; low coverage: five reads; high coverage: 99.5 percentile of each sample's total coverage. Sites with missing data and those failing to pass quality thresholds were replaced with Ns in the matrix. Five individuals, each with fewer than 2.5 million bases passing the initial filtering step, were removed from the analysis. The final number of individuals passing this filtering step and included in downstream

analyses were: *A. alcalica* n=38; *A. latilabris*: n=19; *A. ndalalani*: n=15; *A. grahami*: n=12; *O. amphimelas*: n=7. Full details and sequence quality measures by individual sample are given in Appendix Table 3A.1. The filtered species vcf files were merged to form a single alignment file. Further filtering included imposition of maximum levels of missing data and a minimum allele frequency threshold; these parameters differed by analysis, so the specific values for each analysis are described below and in Table 3.2.

For reads that did not align to the reference genome, unmapped reads were extracted, read1 was clustered using Stacks with a minimum threshold for clustering of 70 out of 96 individuals, and corresponding read2 assembled *de novo* using IDBA-UD (Peng *et al*. 2012). A consensus sequence was then generated from the resultant reads, and any reads not mapping back to the consensus were discarded. All downstream processing and SNP-calling for the unmapped reads followed that described above for the reference-aligned reads, using the consensus sequence as the pseudo-reference from which to make genotype calls in the GATK. The bioinformatics pipeline is summarised in Figure. 3.2.

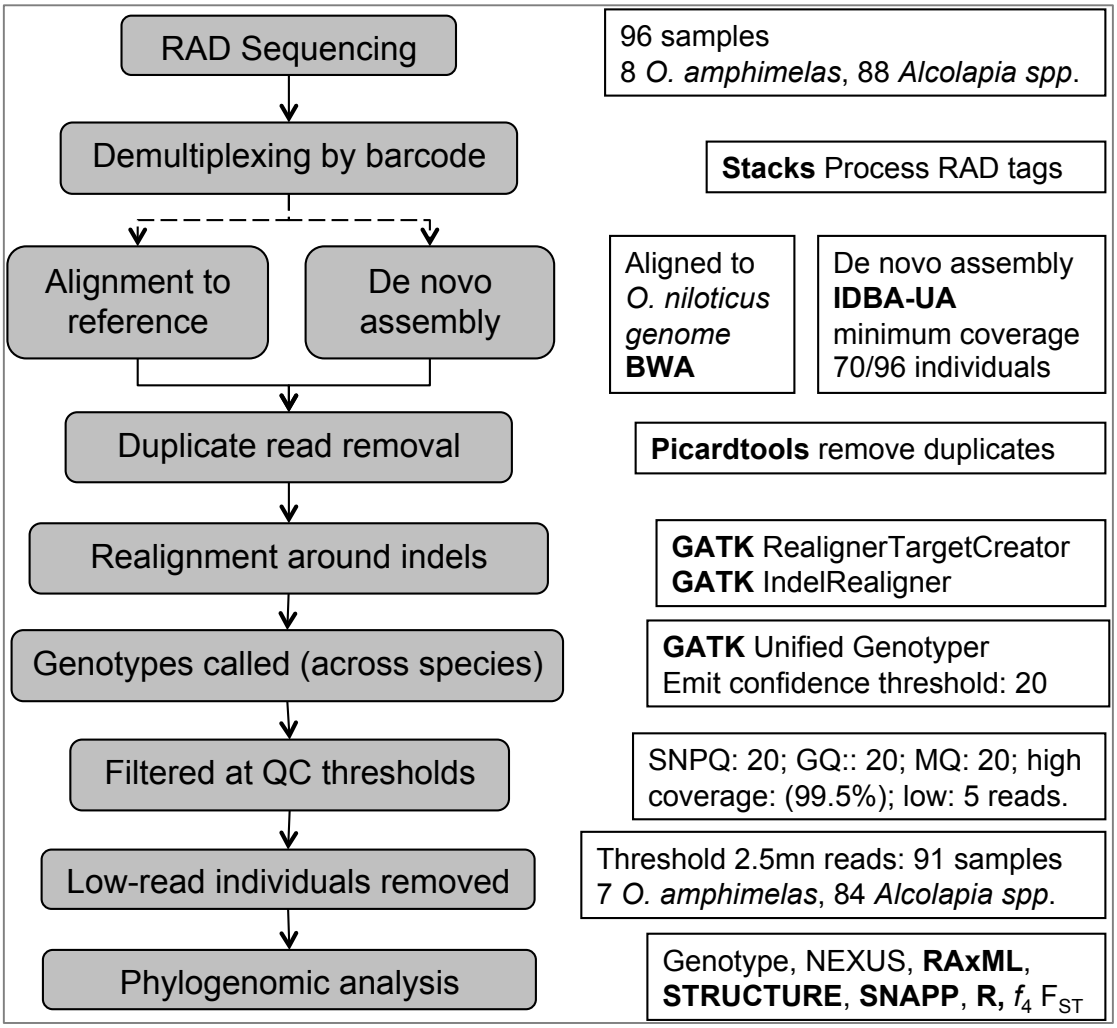


Figure 3.2. Bioinformatic processing pipeline for RAD sequence data.

Table 3.2. Data subsets and respective analyses conducted on RAD data

	Description	Sequences	Sites (n)	Var.	Analyses
A	Full alignment of all mapped, filtered sites with data for >1 individual	n=92 (ref. *O. niloticus*, 7 *O. amphimelas*, 84 *Alcolapia*)	28,560,698	1.91	p-distance
B	Full alignment of all mapped, quality filtered sites, reduced taxa	n=25 (ref. *O. niloticus*, 4 *O. amphimelas*, 20 *Alcolapia*)	26,135,098	1.71	RAxML GTRGAMMA model
C	Full alignment excluding reads not mapped to assigned linkage groups	n=92 (ref. *O. niloticus*, 7 *O. amphimelas*, 84 *Alcolapia*)	22,249,264	1.82	LD, f_4 F_{ST} (EggLib method)
D	Variable sites (SNPs) across full alignment	n=92 (ref. *O. niloticus*, 7 *O. amphimelas*, 84 *Alcolapia*)	544,916	100	RAxML GTRGAMMA model
E	SNPs across full alignment, excluding ambiguous bases	n=92 (ref. *O. niloticus*, 7 *O. amphimelas*, 84 *Alcolapia*)	366,016	100	RAxML GTRGAMMA model, corrected for ascertainment bias
F	Biallelic unlinked SNPs 12 populations (n=4 per population), no missing data	n=48 (38 *A. alcalica*, 12 *A. grahami*, 19 *A. latilabris*, 15 *A. ndalalani*)	1,266	100	SNAPP
G	Full alignment, *de novo* assembled, filtered sites	n=91 (7 *O. amphimelas*, 84 *Alcolapia*)	436,839	1.33	RAxML GTRGAMMA model
H	SNPs from *de novo* assembled alignment	n=91 (7 *O. amphimelas*, 84 *Alcolapia*)	5,832	100	RAxML GTRGAMMA model
J	SNPs from *de novo*, no ambiguous bases	n=91 (7 *O. amphimelas*, 84 *Alcolapia*)	1,898	100	RAxML GTRGAMMA model, _ASC model
K	Combined mapped and *de novo* reads, SNPs	n=91 (7 *O. amphimelas*, 84 *Alcolapia*)	550,748	100	RAxML GTRGAMMA model
L	Combined mapped and *de novo* reads, SNPs, no ambiguous bases	n=91 (7 *O. amphimelas*, 84 *Alcolapia*)	182,909	100	RAxML GTRGAMMA model, _ASC model
M	Biallelic SNPs across *Alcolapia*	n=84 (38 *A. alcalica*, 12 *A. grahami*, 19 *A. latilabris*, 15 *A. ndalalani*)	246,336	100	STRUCTURE, SplitsTree
N	Biallelic unlinked SNPs across *Alcolapia*	n=84 (38 *A. alcalica*, 12 *A. grahami*, 19 *A. latilabris*, 15 *A. ndalalani*)	2,297	100	STRUCTURE
P	*A. alcalica* biallelic SNPs	n=38 *A. alcalica*	173,964	100	STRUCTURE
Q	*A. alcalica* unlinked SNPs	n=38 *A. alcalica*	2,222	100	STRUCTURE
R	Biallelic SNPs, max missing data 10%	n=84 (38 *A. alcalica*, 12 *A. grahami*, 19 *A. latilabris*, 15 *A. ndalalani*)	31,555	100	F_{ST} (Arlequin)
X	Biallelic unlinked SNPs, sympatric *Alcolapia*	n=24 (8 *A. alcalica*, 8 *A. latilabris*, 8 *A. ndalalani*)	2,227	100	STRUCTURE

Var.: Percentage sites variable in full alignments (i.e., including full sequence data).

Estimation of the extent of linkage disequilibrium

As several downstream analyses required the use of unlinked SNPs, and linkage disequilibrium (LD) has not previously been investigated in *Alcolapia*, LD was estimated for each species using the R package snpstats (Clayton & Leung 2007), with R scripts modified from Martin *et al.* (2013), using the reference-aligned dataset. LD was estimated between pairs of SNPs on each linkage group, and the correlation coefficient r^2 of all pairs was averaged within specified distance bins. Background LD was estimated by calculating r^2 between all pairs of SNPs on different linkage groups. Any scaffolds that were not assigned to specific linkage groups of the reference genome (i.e., those positions within the genome designated as unknown scaffolds, which accounted for 22% of the full alignment; dataset C, Table 3.2) were not included in these calculations. Varying the values of the minimum allele frequency (MAF) threshold and the allowed limit of missing data, as well as the number of individuals included per species, had large effects on the dropoff rate of LD, so conservative thresholds were used in the final analysis. LD was estimated for each species separately using only SNPs from the aligned dataset with a MAF of 0.2 and for sites at which there was data for a minimum of 80% of individuals. The LD dropoff plots are presented for analysis using 12 individuals per species for all the *Alcolapia* species. Additionally, for *A. grahami* analysis was for n=8, including only the Lake Magadi samples, and excluding those from the introduced population in Lake Nakuru. Given the apparent divergence between lakes for *O. amphimelas*, only the analysis of LD dropoff for *O. amphimelas* samples from Lake Eyasi (n=4) is presented; although including the additional samples from Lake Manyara had little impact on the result (data not shown).

Phylogenomic inference

Maximum likelihood (ML) phylogenetic inference was conducted using RAxML (Stamatakis 2014) implementing a rapid bootstrap search on all datasets (Table 3.2), with *O. niloticus* (for mapped reads) or *O. amphimelas* (for *de novo* assembled reads) specified as the outgroup for 100 bootstrap replicates. Most analyses were performed within the CIPRES Science Gateway V. 3.3 (Miller et al. 2010) using the RAxML-HPC2 version on XSEDE, with the default models using GTRCAT for the bootstrapping phase and GTRGAMMA for the final tree inference. A reduced taxon dataset was used for ML analysis of the full alignment (including invariant sites), and selected the sample(s) from each population with the highest sequencing quality, for a total of 25 taxa. This full-alignment dataset was analysed using the SSE

PTHREADS version of RAxML 8. For SNP-only datasets (i.e., alignments including no invariant sites) the RAxML GTRGAMMA model was run with and without the ASC correction for ascertainment bias that may be more appropriate for SNP alignments containing no constant sites (RAxML 8 Manual); however, as the ASC model runs only on variant sites and does not consider ambiguous bases variable if the base could be the same as determined bases at that site, this necessitated using a different dataset (excluding ambiguous bases). Given the difficulty in estimating ingroup relationships within this dataset (see Results), a Neighbour-Net algorithm (Bryant 2004) was also used, based on uncorrected p-distances implemented in SplitsTree 4.13.1 (Huson & Bryant 2006) and drawn using the equal-angle algorithm.

A species tree was estimated using the Bayesian software program SNAPP v 1.1.4, (Bryant *et al.* 2012) as an add-on package to BEAST v 2.1.3 (Bouckaert *et al*. 2014). Owing to the prohibitive increase in computational requirements with increasing taxa and individual number, a reduced dataset was used. Population membership was pre-defined based on taxonomic species by sampling site, and included sites where Lake Natron species occurred sympatrically and full sample sizes were available (sites 05 and 11), sites exhibiting *A. alcalica* morphs (site 15), the northernmost Natron sampling site (site 19), and only Lake Magadi *A. grahami* sites (18 and 21). The dataset comprised all biallelic SNPs across 44 *Alcolapia* samples (four samples for each of 11 populations) and four *O. amphimelas* samples, with a minimum distance between SNPs of 500 kb on each linkage group (to ensure SNPs were unlinked) and any sites for which data was not available for all samples were removed, leaving a total of 1,266 SNPs. Backward and forward (*u* and *v*) mutation rates were estimated from the data using equation 8.4.1 from (Drummond & Bouckaert 2014). The analysis used a gamma prior with parameters to account for small population sizes (α=2, β=2,000, with θ=0.001) and each analysis was run for 7 million generations, discarding the first 10% as burn-in. Runs were checked for convergence using Tracer v 1.5 software (Rambaut & Drummond 2007), ensuring that each reached an effective sample size (ESS) >200. Resultant tree sets were visualised using DensiTree (Bouckaert 2010).

Population genomic analyses

Population clustering of *Alcolapia* populations were assessed using STRUCTURE v 2.3.4 (Pritchard *et al*. 2000). As the underlying clustering algorithm of STRUCTURE assumes markers are unlinked loci, biallelic SNPs were used with a minimum

distance imposed of 500kb between SNPs on the same linkage group, resulting in a dataset of 2,297 SNPs across the 84 *Alcolapia* samples (Table 3.2). For comparison these analyses were also run on the full dataset without accounting for linkage disequilibrium. Given the very recent divergence of *Alcolapia* species, an analysis using the LOCPRIOR model (Hubisz *et al.* 2009) was also employed, using taxonomic species as a prior, which can provide more accurate inference of population structure when the signal is too weak for standard models to detect. For species priors the four described species were specified, and additionally a 5th category was included for two samples that were originally identified as *A. alcalica*, but after further inspection were re-classified as *A.* aff. *ndalalani* (see Results).

The allele frequency parameter (λ) was estimated using an initial run of K=1 with 50,000 burnin and 100,000 further iterations, giving a value of λ=0.5252. This value was set in subsequent runs of 5 iterations at each value K=1–12 with no prior population information, and 50,000 burnin/100,000 further iterations. Analyses were run with all different model parameters independently (total of four separate analyses: admixture model / allele frequencies correlated (default settings); admixture model / independent allele frequencies; no admixture / allele frequencies correlated; no admixture / allele frequencies independent). STRUCTURE output was compiled and averaged using Structure Harvester (Earl & vonHoldt 2011) to conduct the Evanno method (Evanno *et al*. 2005), and run permutations were clustered using CLUMPP v 1.1.2 (Jakobsson & Rosenberg 2007). Finally, clustered output was visualised using Distruct v 1.1 (Rosenberg 2004). In addition to the STRUCTURE analysis of the entire *Alcolapia* dataset, analyses were also conducted on subsets of the data. As the full-dataset STRUCTURE runs indicated differences in cluster membership between allopatric and sympatric sites, analysis was also conducted on a dataset containing only individuals occurring sympatrically in Lake Natron (sites 005 and 011) and analyses were run separately and combined, with a minimum of 500kb between SNPs and the following parameters: site 05: 12 individuals; 2,180 SNPs; λ=0.8; K=1-8; site 11: 12 individuals; 2,160 SNPs; λ=0.8; K=1-8; site 05 and 11 combined: 24 individuals; 2,227 SNPs; λ=1.1; K=1-16. As the *Alcolapia* dataset indicated separation within the *A. alcalica* samples (see Results section), analysis was repeated on a separate dataset containing only *A. alcalica* samples (38 samples, 2,222 SNPs), with an obtained value of λ=0.7184, for five independent runs at each K value of K=1–8.

Pairwise comparisons

Uncorrected pairwise p-distances between samples were calculated in the package ape (Paradis *et al*. 2004) using R v 2.15.2 (R Development Core Team 2012). Calculations of pairwise F_{ST} to test genomic differentiation between populations were conducted in the EggLib Python module (De Mita & Siol 2012). Whole-dataset F_{ST} values were estimated by averaging over non-overlapping windows of 100kb, which has been shown to provide accurate estimates for small sample numbers (Nadeau *et al.* 2012). Any windows returning negative values for F_{ST} were removed before averaging. F_{ST} was also calculated in Arlequin 3.1.5.2 (Excoffier *et al.* 2005), accounting for differences in sample size between populations, and tested significance using 10,100 permutations.

As recently diverged species are likely to continue to exchange genes through interspecific hybridisation (e.g., Nosil *et al.* 2009), the extent of ongoing gene flow between species was examined using the f_4 four-population test for admixture (Reich *et al.* 2012), which is based on the fact that genetic drift should be uncorrelated in unadmixed populations. The f_4 test was used rather than tests of phylogenetic discordance (e.g., ABBA-BABA tests; Durand *et al*. 2011), which may be confounded by the presence of gene flow between sympatric taxa. The test was conducted between each of the Natron species at three sympatric sites with varying geographic distance (sites 05, 12, 17). The f_4 statistic mean and variance was calculated with a block jack-knifing approach (block size of 500kb, as identified by LD estimates) using modified python scripts adapted from (Martin *et al.* 2013; Dryad Digital Repository. doi:10.5061/dryad.dk712). The F_{ST} and f_4 statistics were calculated using only reads that aligned to assigned linkage groups within the reference genome. Following phylogenomic analysis and an unexpected placement within the resulting phylogeny (see Results section), two samples were found to represent intermediate morphology between species, and were excluded from population comparisons for F_{ST} and the f_4 test.

Isolation by distance

To investigate the hypothesis of panmixia within the lakes, Mantel tests were used to test for covariation of genetic distance with geographic distance between populations. Mantel tests were conducted using the ade4 package (Dray & Dufour 2007) to test matrix covariation in R 2.15.1 for population pairwise F_{ST} comparisons vs. geographic distance between sampling sites.

Results

Generation of a genome-wide SNP dataset using RAD Sequencing

A total of 83.6 Gb of sequence was produced, of which 89% successfully mapped to the *O. niloticus* reference genome in the alignment stage. Mapping, duplication, and filtering statistics per individual sample are provided in supplementary information in Appendix Table 3A.1. Mapping rates were generally consistent across species, although slightly lower for *A. grahami* than the other species (mean mapping rates: *O. amphimelas*: 89.9%; *A. alcalica*: 88.8%; *A. latilabris*: 88.7%; *A. ndalalani*: 88.8%, *A. grahami*: 86.7%). Following removal of duplicates and poorly paired reads, the dataset comprised 272,360,470 paired reads. From these filtered reads, a total of 1,229,734,617 calls were generated using the GATK UnifiedGenotyper (mean: 12,809,736; range 313,244–21,316,544 calls). Five samples with fewer than 2.5 million calls were removed from further analysis, resulting in a dataset of 91 samples (84 *Alcolapia* species and 7 *O. amphimelas*). Quality filtering (see *Methods* for specific parameters) resulted in a final dataset of 28,560,699 retained sites, representing ~3% of the reference genome. The final dataset comprised 544,916 (1.91%) sites that were variable across the whole alignment including *O. niloticus*, of which 238,203 sites were variable across *Alcolapia* samples and 164,014 sites were variable across *A. alcalica* samples alone.

A separate dataset was compiled from reads that did not map to the *O. niloticus* reference genome, as such reads are likely to lie in more divergent regions of the genome. After quality filtering as for the mapped reads, the dataset contained a total of 436,839 sites, of which 5,832 (1.34%) were variable across the 91-sample dataset. The combined dataset of mapped and unmapped reads resulted in a total alignment of 550,748 variable sites (fewer than that from the mapped dataset alone as the *O. niloticus* reference sites were removed before combining). The final datasets used in different analyses for phylogenomic inference and population genomic analysis are detailed in Table 3.2.

Linkage disequilibrium

Values of r^2 were estimated for pairs of SNPs on the same linkage group, and plotted against distance between SNPs. LD decreased with distance (Figure 3.3), and reached background level (mean LD between SNPs on different linkage groups) at 100-500 kb. This distance was smaller in *A. alcalica* and *A. grahami* than

in *A. latilabris* and *A. ndalalani*, suggesting larger population sizes in these species. For all downstream analyses that required unlinked SNPs, a minimum distance between SNPs of 500 kb was imposed.

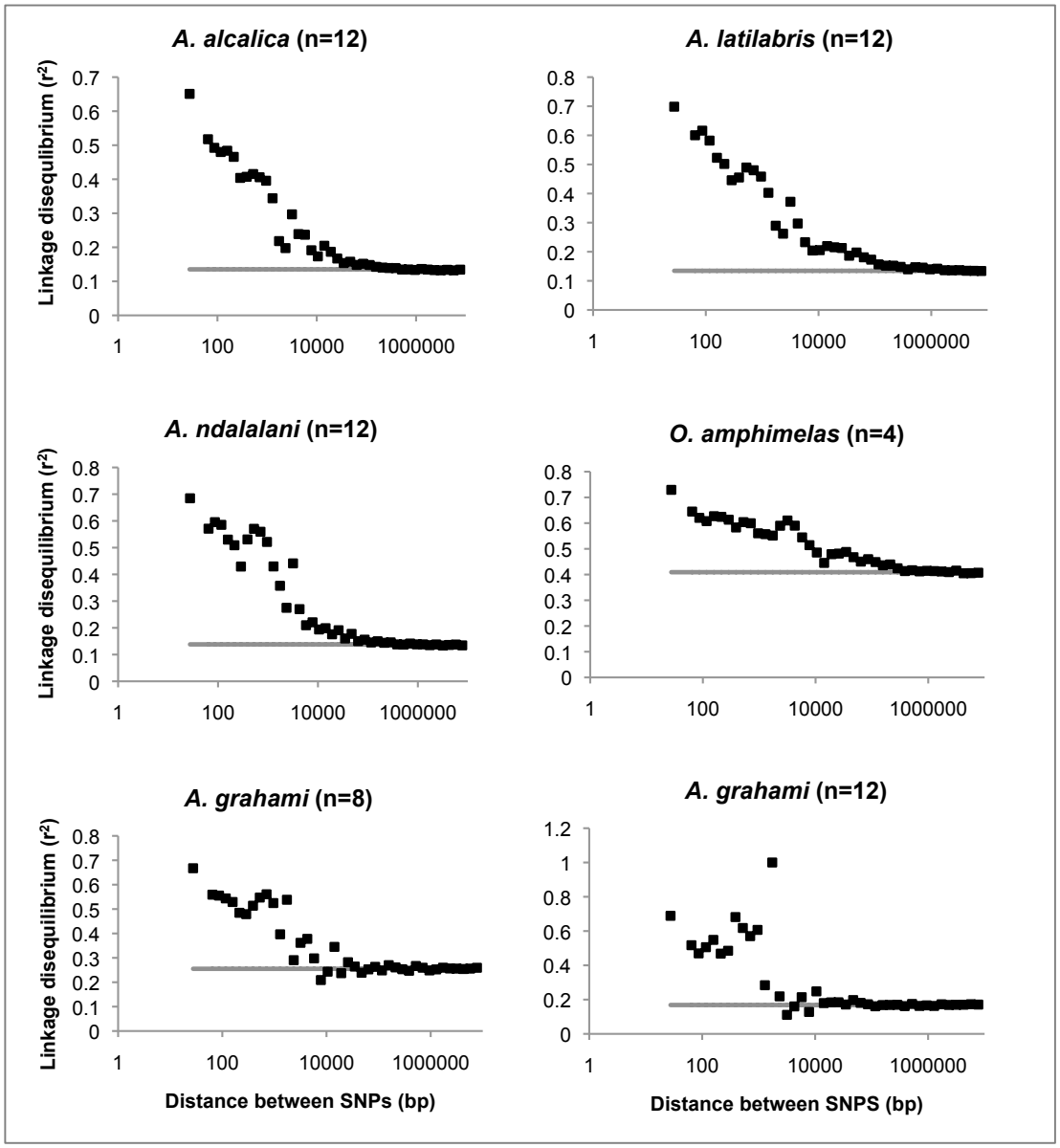


Figure 3.3. Linkage disequilibrium dropoff with distance by species.

Grey line in each plot indicates background LD level.

Phylogenomic inference

The ML phylogeny of the full mapped-read alignment reduced-taxa dataset (dataset B; 26 million bp; n=25) provides maximum support for a clade composed of *A. grahami* individuals as sister to the clade comprising Lake Natron *Alcolapia* (Figures 3.4A and 3.4B). However, there is weak support for the monophyly of the Lake

Natron *Alcolapia* species, with *A. alcalica* from the northern populations sister to all species from the sympatric southern populations (i.e., *A. alcalica*, *A. ndalalani*, *A. latilabris*). Short branch lengths within the ingroup relative to the outgroup indicate very low genomic differentiation. In contrast there is maximum support for the separation of *O. amphimelas* (outgroup) populations between Lake Eyasi and Lake Manyara, and branch lengths between these populations are considerably longer than across the entire *Alcolapia* radiation.

Considering the *Alcolapia* populations in more detail, ML analysis was conducted on the full taxon dataset, but included only SNPs in the analysis (datasets D & E; Figures 3.4C and Appendix Figure 3A.3). The SNP-only phylogenies indicate that the Lake Natron do not form a clade, but support for this relationship is weak. The two species restricted to the southern lagoons, *A. latilabris* and *A. ndalalani,* form well-supported sister clades to the exclusion of individuals from site 17, whilst the ubiquitous Lake Natron species *A. alcalica*, comprises two separate clusters separated by sampling locality of northern and southern sites.

Notably, most taxa from site 17 comprise a clade, rather than clustering by species, and are sister to the northern *A. alcalica* clade. The genomic signal from individuals at site 17 does not reflect existing species designations. While all four *A. ndalalani*, two *A. latilabris* samples and a single *A. alcalica* specimen form a separate, well-supported clade, other site 17 individuals occur elsewhere within the tree and two *A. alcalica* samples nest within *A. ndalalani*. Voucher material was re-examined for all site 17 specimens, and were confirmed with original morphology-based species identification for all samples except the two *A. alcalica* individuals nesting within *A. ndalalani* (017-844-AA and 017-846-AA) that exhibited an intermediate form, and were thereafter designated *A.* aff. *ndalalani*.

The ML tree also generally exhibits populations (sampling sites) that cluster together within these clades. Given the uncertainty of some branch placements and low bootstrap support for *A. alcalica* nodes, this dataset was also visualised as a phylogenetic network based on the *Alcolapia* samples only (246,366 SNPs, dataset M; Figure 3.4D). The network again highlights the close relationships of all Lake Natron species, and while the *A. grahami* samples cluster with *A. alcalica*, this species is well separated from the Natron species (Figure 3.4D). Across all the phylogeny and network analyses, there was no discernible genomic differentiation between intraspecific morphs.

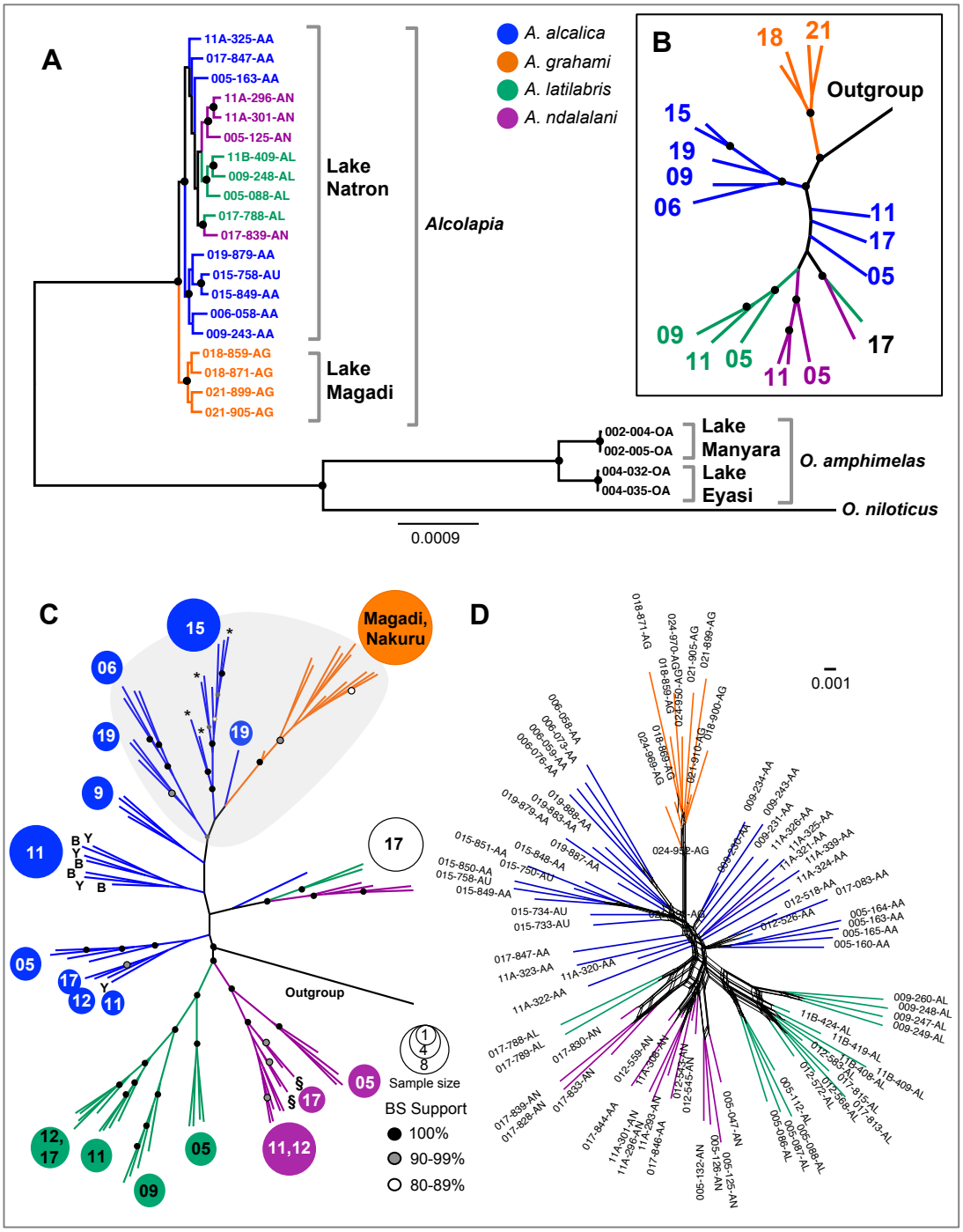


Figure 3.4. Phylogenomic analysis of reference-aligned RAD sequence data.

Figure 3.4. Legend for figure on previous page.

Phylogenomic analysis of reference-aligned RAD sequence data.

A) Maximum likelihood phylogeny (RAxML) for the reduced-taxon dataset (n=25), full quality-filtered alignment (dataset B); B) Radial tree layout for tree in panel A; C) ML phylogeny of the full taxon dataset (n=92) alignment of variable sites only (dataset D); D) Phylogenetic network (NeighbourNet) of ingroup taxa (dataset M; 84 samples). B-C: Numbers at tips indicate the sampling location (population) of individuals in each clade; branch length to outgroup has been truncated for clarity. C: Northern Lake Natron sites and Magadi/Nakuru sites are shaded in grey. *indicates *A. alcalica* upturned-mouth morph individuals from site 15. §indicates possible hybrids that displayed intermediate morphology between *A. alcalica* and *A. ndalalani* from site 17. B and Y at tips indicate, respectively, blue or yellow *A. alcalica* morphs found at site 11.

Maximum likelihood trees generated from the *de novo* assembly datasets (Table 3.2; datasets G, H and J) did not resolve species within *Alcolapia*, and individuals did not cluster by site or species. Given the lack of resolution for almost all nodes, these phylogenies are presented as majority rule (50%) consensus trees, and almost the entire *Alcolapia* forms a polytomy (Appendix Figure 3A.4, A-C). However, the 100% bootstrap support for *O. amphimelas* nodes was maintained – both for the node separating *O. amphimelas* from *Alcolapia*, and the node separating *O. amphimelas* specimens between Lakes Manyara and Eyasi. When the *de novo* datasets were combined with the mapped-read dataset (datasets K and L) ML analysis exhibited similar topology to the alignment of mapped reads alone and grouped by species, but with lower bootstrap support (Appendix Figure 3A.4 D and E).

A potential difficulty in the interpretation of the phylogenomic analysis is the placement of the root in the different tree topologies. As the outgroup (*O. amphimelas*) is comparatively distant to the ingroup that has diverged rapidly, it is difficult to place the root accurately, which in turn can influence the ingroup topology (Kirchberger *et al.* 2014). However, ML analysis of ingroup data excluding the outgroup resulted in the same overall tree topology, with taxa clustering by population within species, suggesting that the outgroup does not affect ingroup relationships (Appendix Figure 3A.5).

Similar to ML analysis of the full alignment (dataset B), the SNAPP species tree (Figure 3.5) also places *A. grahami* as sister to a clade composed of all Lake Natron terminals, but reveals a deeper divergence between this taxon and those from Lake Natron. Furthermore, the species analysis demonstrated long branch lengths to the

outgroup taxa, but very close relationships within *Alcolapia* species from southern Natron populations.

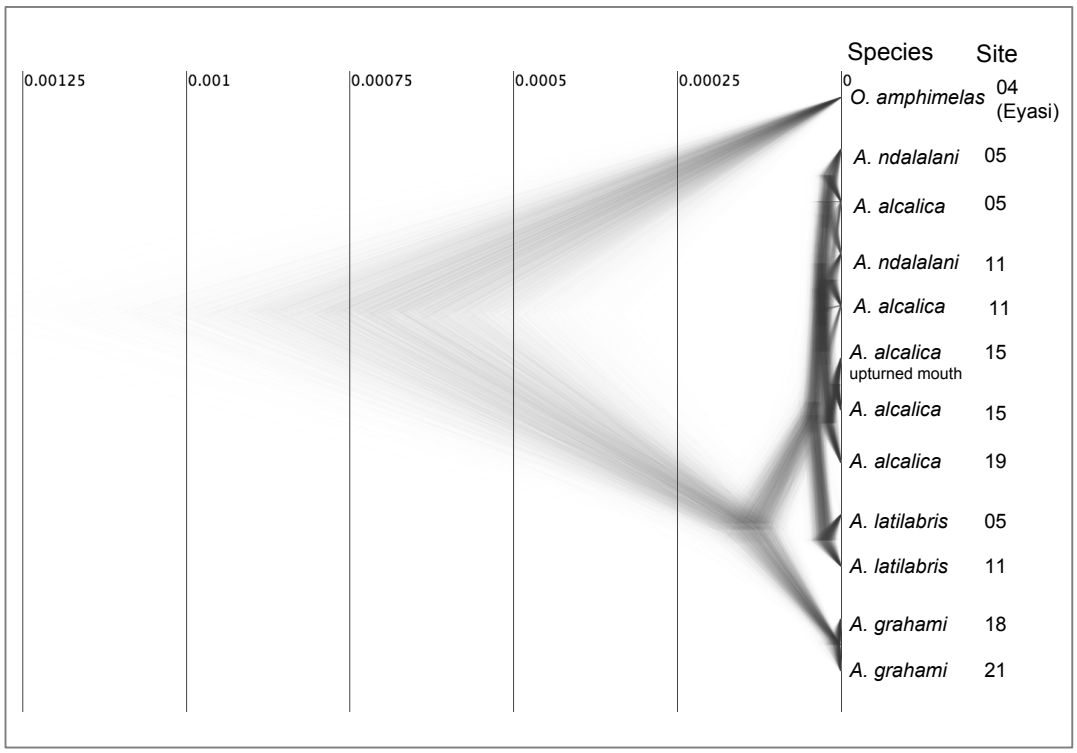


Figure 3.5. Species tree generated by SNAPP analysis (selected populations). Analysis for a reduced-taxon dataset of 4 individuals from each of 12 populations, with unlinked biallelic SNPs (dataset F; n=48; 1,266 SNPs).

Population clustering and admixture

STRUCTURE analysis of *Alcolapia* unlinked SNPs (dataset N) gave the highest likelihood scores for the admixture and correlated allele frequencies models. Using these models, lnP(K) gave an optimum of K=4 and the Evanno method exhibited a modest peak at K=3 (both K values visualised in Figure 3.6). Running the analysis with the LOCPRIOR model and species prior information gave a clear optima of K=3, as did running the analysis across the full dataset (not accounting for LD; dataset M; Appendix Figure 3A.6). The cluster membership at K=3 and K=4 reflects the differentiation observed in the ML tree, with shared cluster membership between species in the sympatric southern Lake Natron populations, but with the allopatric northern populations of *A. alcalica* showing strong probability of membership of a single cluster. Furthermore, *A. grahami* is assigned to a distinctly separate cluster from all other individuals, with no mixing.

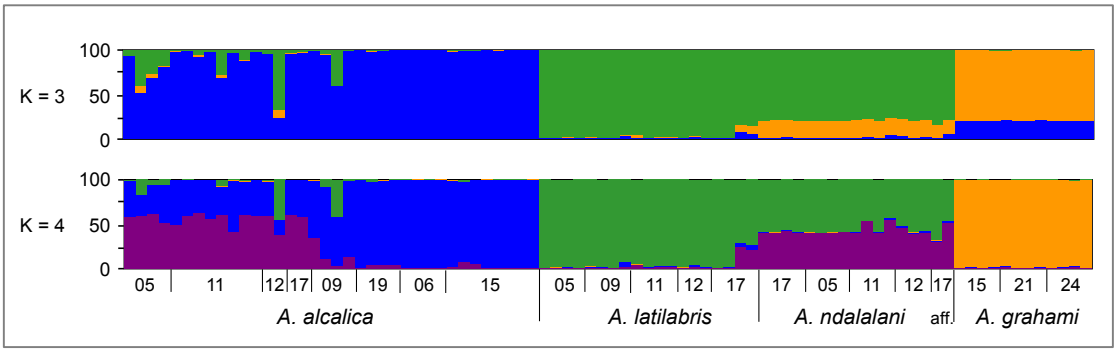


Figure 3.6. STRUCTURE analysis of *Alcolapia* populations. Analysis of the dataset accounting for linkage disequilibrium, imposing a minimum distance of 500kb between SNPs (dataset N; 84 samples; 2,297 unlinked biallelic SNPs), using the admixture and correlated allele frequency models with STRUCTURE. Clusters are visualised for the most likely K values (K=3-4). Results for alternative models and the full SNP alignment are provided in Appendix Figures 3A.5-3A.7.

When the Lake Natron sympatric sites were analysed separately (Appendix Figure 3A.7), both site 05 and site 011 had an lnP(K) optimum of K=1 and an Evanno method optimum at K=2, but the cluster membership differed between sites, with the site 05 analysis indicating most likely membership to a single cluster across all species, while site 11 indicated different cluster membership proportions by species. When analysed jointly, optima were seen at K=2 and K=4, with the latter K value indicating differing cluster membership between species, and between populations with *A. alcalica* (Appendix Figure 3A.7). Given the separation of northern and southern *A. alcalica* populations indicated by both the phylogenetic inference and STRUCTURE analyses, STRUCTURE analysis was also conducted on a dataset including only the *A. alcalica* samples (dataset Q) in order to disentangle population clustering within this species. Replicate runs gave an optimum of K=1, suggesting that populations within *A. alcalica* are not sufficiently diverged to detect population structure within the species at this scale (Appendix Figure 3A.8).

As the STRUCTURE results indicated high levels of admixture between species, four-population tests specifically designed to test admixture were conducted within the Natron species. The f_4 population test comparisons were all significant, indicating admixture between all species pair combinations: *A. alcalica-A. ndalalani*, *A. alcalica-A. latilabris* and *A. latilabris-A. ndalalani* (Table 3.3). Furthermore, the test Z-score for the comparison of *A. latilabris* and *A. ndalalani* at site 17 was substantially higher than for other comparisons, suggesting increased gene flow between species at this site.

Table 3.3. Four-population test for recent gene flow.

A	B	C	D	f_4 ± SEM	Z-score	P-value
05_Aa	12_Aa	05_An	12_An	0.011 ± 0.002	6.369	$P=1.9 \times 10^{-10}$
05_Aa	12_Aa	05_Al	12_Al	0.012 ± 0.002	6.830	$P=8.5 \times 10^{-12}$
05_An	12_An	05_Al	12_Al	0.012 ± 0.002	7.400	$P=1.4 \times 10^{-13}$
05_Aa	17_Aa	05_An	17_An	0.010 ± 0.001	9.782	$P=1.3 \times 10^{-22}$
05_Aa	17_Aa	05_Al	17_Al	0.010 ± 0.001	10.672	$P=1.4 \times 10^{-26}$
05_An	17_An	05_Al	17_Al	0.021 ± 0.001	21.522	$P=9.8 \times 10^{-103}$

The f_4 statistical test between 2 pairs of populations (A,B;C,D) - A significant Z-score indicates gene flow, with positive values implying flow between populations A-C and/or B-D. Aa: *A. alcalicus*; Al: *A. latilabris*; An: *A. ndalalani*. The number in each population name refers to the sampling location as numbered in Figure 3.1.

Inter-population F_{ST} values exhibited a similar pattern when calculated from the entire dataset using EggLib or when using a reduced dataset (maximum 10% missing data) in Arlequin. Pairwise population F_{ST} values revealed higher levels of differentiation between more distant populations and between described species (Table 3.4). As expected, *A. grahami* populations exhibited the highest F_{ST} values when compared to Lake Natron populations, indicating population structuring between the two lakes, with lower pairwise F_{ST} values between *A. grahami* and the most northern *A. alcalica* population (site 19). Individuals from site 17 in Lake Natron exhibited the lowest interpopulation and interspecific values, which correlates with relationships as indicated by the ML tree and suggests very recent or contemporary hybridisation occurring at this site. The range of intraspecific F_{ST} values between populations was similar across Lake Natron species (*A. alcalica*: 0.001-0.113; *A. latilabris*: 0.030-0.146; *A. ndalalani*: 0.000-0.143), but substantially lower in Lake Magadi *(A. grahami*: 0.000-0.014). However, after correcting for multiple tests (Bonferroni correction), none of these population comparisons was significant.

Table 3.4. Population pairwise F_{ST}.

Above diagonal: Population pairwise F_{ST} calculated for entire dataset (EggLib). Below diagonal: F_{ST} calculated in Arlequin accounting for sample size.

		A. alcalica									*A. latilabris*					*A. ndalalani*				*A. grahami*		
Site:		**005**	**006**	**009**	**011**	**012**	**015**	**015***	**017**	**019**	**005**	**009**	**011**	**012**	**017**	**005**	**11A**	**012**	**017**	**018**	**021**	**024**
AA	005	-	0.093	0.051	0.033	0.033	0.076	0.039	0.103	0.058	0.077	0.104	0.101	0.110	0.063	0.082	0.086	0.076	0.102	0.114	0.103	0.109
	006	0.075	-	0.055	0.052	0.131	0.074	0.080	0.096	0.042	0.135	0.136	0.156	0.148	0.102	0.145	0.134	0.108	0.135	0.127	0.113	0.122
	009	0.038	0.039	-	0.054	0.084	0.046	0.037	0.069	0.028	0.086	0.072	0.092	0.094	0.058	0.085	0.082	0.065	0.086	0.097	0.083	0.087
	011	0.049	0.058	0.011	-	0.074	0.067	0.025	0.053	0.047	0.055	0.057	0.063	0.083	0.045	0.050	0.047	0.051	0.052	0.115	0.106	0.099
	012	0.056	0.084	0.017	0.016	-	0.118	0.069	0.155	0.093	0.151	0.154	0.167	0.151	0.093	0.151	0.152	0.109	0.150	0.123	0.136	0.150
	015	0.099	0.078	0.068	0.074	0.113	-	0.063	0.031	0.042	0.109	0.115	0.130	0.129	0.081	0.119	0.109	0.097	0.108	0.102	0.090	0.093
	015*	0.072	0.058	0.044	0.056	0.083	0.001	-	0.083	0.047	0.070	0.084	0.080	0.082	0.037	0.055	0.034	0.030	0.062	0.107	0.095	0.100
	017	0.029	0.053	-0.003	0.001	0.004	0.074	0.034	-	0.061	0.145	0.145	0.160	0.160	0.103	0.147	0.131	0.125	0.128	0.126	0.119	0.126
	019	0.042	0.027	0.001	0.020	0.038	0.057	0.038	0.022	-	0.102	0.099	0.119	0.112	0.068	0.098	0.095	0.072	0.099	0.085	0.075	0.079
AL	005	0.063	0.117	0.072	0.083	0.096	0.126	0.105	0.059	0.087	-	0.108	0.082	0.103	0.061	0.103	0.112	0.110	0.129	0.156	0.141	0.143
	009	0.146	0.158	0.085	0.127	0.147	0.174	0.151	0.135	0.116	0.132	-	0.070	0.100	0.063	0.136	0.126	0.119	0.140	0.158	0.149	0.152
	011	0.102	0.124	0.078	0.084	0.104	0.138	0.115	0.075	0.091	0.080	0.082	-	0.058	0.045	0.130	0.105	0.115	0.140	0.159	0.137	0.151
	012	0.132	0.153	0.101	0.115	0.118	0.176	0.153	0.110	0.121	0.106	0.146	0.026	-	0.045	0.152	0.138	0.111	0.152	0.154	0.143	0.151
	017	0.053	0.137	0.042	0.043	0.050	0.100	0.072	0.018	0.039	0.067	0.093	0.033	0.030	-	0.092	0.079	0.063	0.068	0.122	0.110	0.113
ANi	005	0.090	0.126	0.071	0.076	0.109	0.154	0.120	0.069	0.081	0.108	0.191	0.140	0.159	0.099	-	0.074	0.092	0.115	0.157	0.144	0.152
	011	0.113	0.137	0.077	0.062	0.072	0.157	0.134	0.068	0.089	0.130	0.163	0.119	0.151	0.075	0.122	-	0.046	0.086	0.149	0.130	0.133
	012	0.072	0.101	0.047	0.031	0.065	0.126	0.105	0.037	0.058	0.105	0.150	0.088	0.116	0.041	0.092	-0.014	-	0.101	0.121	0.119	0.113
	017	0.101	0.119	0.076	0.091	0.119	0.136	0.107	0.065	0.089	0.126	0.203	0.141	0.152	0.063	0.143	0.095	0.092	-	0.149	0.138	0.142
AG	018	0.112	0.123	0.100	0.100	0.119	0.137	0.114	0.075	0.089	0.150	0.185	0.154	0.189	0.108	0.164	0.168	0.141	0.156	-	0.033	0.031
	021	0.088	0.103	0.077	0.090	0.093	0.119	0.094	0.056	0.070	0.126	0.170	0.140	0.170	0.097	0.124	0.150	0.123	0.139	-0.010	-	0.030
	024	0.108	0.109	0.088	0.106	0.115	0.135	0.112	0.068	0.079	0.137	0.172	0.147	0.179	0.101	0.154	0.164	0.132	0.142	0.014	-0.007	-

*indicates *A. alcalica* upturned-mouth morph from site 015. Site 011 *A. alcalica* includes comparisons for the blue morph only (n=4). Site 017 *A. alcalica* population excludes the two samples designated *A.* aff. *ndalalani*. None of the comparisons was significant after Bonferroni correction for multiple testing (α=0.0002).

	<0.03		<0.09		<0.15			>0.18	
	<0.06		<0.12		<0.18				

Phylogenomic covariation with geography

To investigate whether each species exhibits panmixia within the respective lakes, isolation by distance was tested by comparing pairwise population F_{ST} comparisons within species to geographic distance between sampling sites. None of the population comparisons exhibited significant covariation with geographical distance (Table 3.5), indicating that genomic divergence between populations is not correlated with distance.

Table 3.5. Mantel test results.

Mean genetic pairwise distance or population pairwise F_{ST} vs. geographic distance between sampling sites.

	Uncorrected p-distance		**F_{ST}**	
Geographic distance:	**Straight-line**	**Lake-perimeter**	**Straight-line**	**Lake-perimeter**
Population comparisons				
A. alcalica	-0.055	0.118	-0.027	0.013
A. latilabris	0.604	0.647	0.169	0.213
A. ndalalani	0.859*	0.879	0.731	0.735
A. grahami	-0.092	-	0.583	-

Alcolapia phylogenomic differentiation

Across the whole dataset there was very low phylogenomic differentiation, with a mean inter-individual uncorrected p-distance for the *Alcolapia* of 0.020%, which was the same as for that within Lake Natron alone (0.019%), and mean interspecific distances only marginally larger: 0.020-0.026% (Figure 3.7). As analysis focused on reads aligned to the *O. niloticus* reference genome, reads that did not align to the genome were also *de novo*-assembled in case these reads represented regions of the genome that had substantially diverged from *O. niloticus*. However, these reads did not appear any more divergent from the outgroup *O. amphimelas* than aligned reads based on genetic distance (uncorrected p-distance; Figure 3.7).

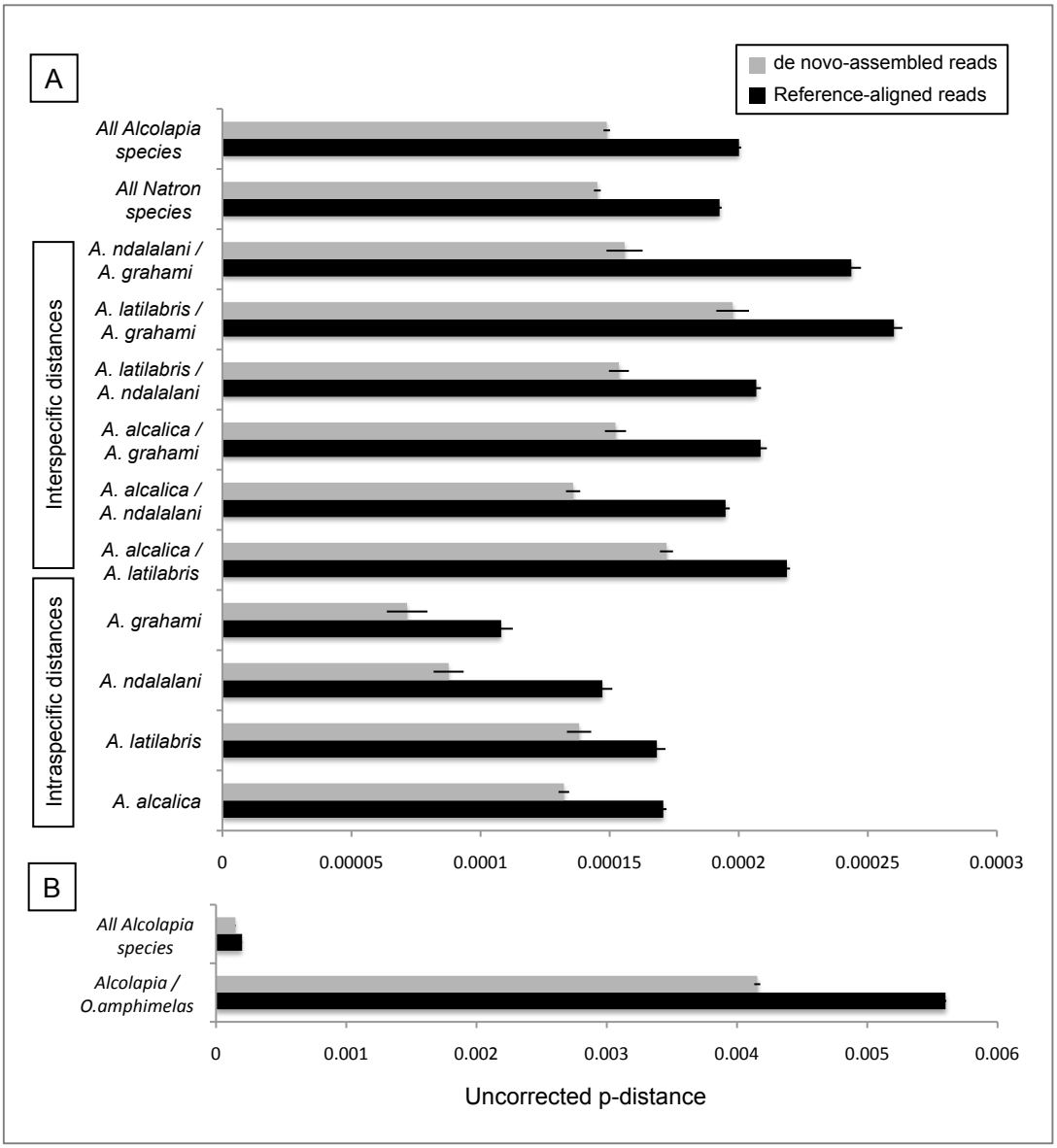


Figure 3.7. Mean inter-specimen uncorrected p-distance.

Comparison of filtered RAD data from reference-aligned and *de novo*-assembled reads. A) Comparisons within *Alcolapia*. B) For comparison, the distance to the outgroup (*O. amphimelas*). Different scale axes are used for each graph (the first entry in each graph is for the same comparison, but at different scales). Error bars are +/- SEM.

Discussion

Despite clear morphological differences and unique physiological adaptations in the soda lake cichlids, previous genetic work has been unable to resolve relationships within the *Alcolapia* radiation. Here, an extensive genomic dataset containing dense sampling of the entire *Alcolapia* is presented and the existing species hypotheses and phylogeny of the entire radiation are addressed.

Alcolapia species relationships

Consistent with described species, the ML tree (Figure 3.4) achieved maximum support for Lake Magadi *A. grahami*, as well as the geographically restricted clades comprising *A. latilabris* and *A. ndalalani*, excluding the anomalous sampling site 17 clade. However, unexpectedly the geographically widely distributed *A. alcalica* is found to be comprised of two distinct groups – clustering by northern and southern localities. This finding is likely to result from geographic isolation of certain *A. alcalica* populations (particularly between the northern and southern lagoons) coupled with gene flow between the southern *A. alcalica* and the other sympatric species. The widespread occurrence of *A. alcalica* across Lake Natron (Figure 3.1) means that several populations are separated by extensive stretches of trona, and there is currently not continuous open water between the northern and southern lagoons. The NeighbourNet network (Figure 3.4D) is congruent with the phylogenomic analyses, showing species-level clustering of *A. grahami*, *A. latilabris* and *A. ndalalani* (excluding site 17 individuals), with higher levels of reticulation between site 17 individuals and the *A. latilabris*/*A. ndalalani* clusters than over the rest of the network. *Alcolapia alcalica* did not form a clade, but clustered by population from the centre of the network, while *A. grahami* appeared the most distinct taxon and had the least reticulation with other groups. The SNAPP analysis (Figure 3.5) produced a species tree topology largely congruent with ML analysis, and confidently resolved *A. grahami* as sister to all the Lake Natron species, with less certainty of Lake Natron species relationships.

In contrast to the close relationships of *Alcolapia* species, considerably more differentiation was found within *O. amphimelas* between the geographically separated Lake Eyasi and Manyara populations than within *Alcolapia* (Figure 3.4), highlighting the possibility of cryptic diversity within the other soda lakes of East Africa.

Population structure

The clusters identified in STRUCTURE analyses at optimum K=3 and K=4 did not definitively separate intraspecific populations within Lake Natron, but do show variable levels of cluster membership by site for *A. alcalica*, with populations on the periphery of the southern Lake Natron lagoon (sites 05, 11, 12,) showing higher levels of cluster membership with sympatric species (Figure 3.6). This suggests that the presence of other species prevents clean clustering by site. Meanwhile those *A. alcalica* populations in isolated lagoons (sites 06, 09, 15, 19) exhibited minimal admixture with other species (Figure 3.6). Differentiation as measured by F_{ST} suggested an effect of geography (Table 3.4); with pairwise population F_{ST} values revealing higher levels of differentiation between more distant populations, and *A. grahami* populations exhibiting the highest F_{ST} values when compared to Lake Natron populations. Despite these differences between sampling sites, there was no correlation between genomic and geographic distances based on Mantel tests (Table 3.5).

The F_{ST} values are slightly higher than, but generally show similar patterns to, those found in a recent Lake Natron study using microsatellites (Zaccara *et al*. 2014), and have similar values to those observed between differentiated cichlid populations in other recent crater lake radiations (Barluenga & Meyer 2004; Elmer *et al.* 2010b). High migration rates between lagoon populations and lack of genetic differentiation have previously been explained by the possibility of heavy rains and flooding increasing permeability of the genetic barrier created by trona crusts (e.g., Zaccara *et al.* 2014). Although even heavy floods may be insufficient to allow panmixia within the lakes, as observations report heavy *Alcolapia* mortality in floodwater between lagoons due to deoxygenation and salinity increase from dissolution of the soda deposits (Coe 1969; Tichy & Seegers 1999; Wilson *et al.* 2004). Previous studies also found morphological and physiological/behavioural differentiation between separate lagoons in Lake Magadi, suggesting local adaptation among populations (Seegers & Tichy 1999; Wilson *et al.* 2004).

Other recent cichlid radiations in which morphs exhibit differences in trophic morphology or colour have shown similarly low levels of genomic differentiation with high levels of phenotypic diversity (e.g., Barluenga & Meyer 2010). One case of divergence between Nicaraguan cichlid Lake Apoyeque morphs, thin- and thick-lipped forms of *Amphilophus cf. citrinellus* are thought to have arisen in only ~100 years (Elmer *et al.* 2010b). In the case of the Midas crater lake cichlids in particular, low levels of genome-wide differentiation between ecologically divergent species and morphs have been shown to be underpinned by selection acting on a few small

genomic regions (Elmer *et al.* 2010a). Outside of the cichlid radiations, a comparably young radiation with similar levels of trophic phenotypic diversity, the <10,000 years old radiation of *Cyprinodon* pupfish on San Salvador Island, exhibited interspecific F_{ST} values of 0.12-0.49 based on RAD data (Martin & Feinstein 2014), whereas the present analysis identified interspecific F_{ST} of only 0.04-0.20 (Table 3.5). However, the interspecific F_{ST} values seen within the current study are within the range of differentiation observed in adaptive divergence in other fish radiations, such as three-spine stickleback populations (0.01-0.13, Hohenlohe *et al.* 2010; 0.03-0.38, Jones *et al.* 2012), Lake Constance *Coregonus* species (0.02-0.08, Vonlanthen *et al.* 2012); whitefish ecotypes (0.001-0.05, Gowell *et al.* 2012); sailfin silversides (0.00-0.21, Schwarzer *et al.* 2008); and fresh and saltwater killifish (0.04-0.40, Kozak *et al.* 2013). Even lower interspecific F_{ST} values have been recorded between fish species differentiated mainly by colour, such as marine hamlets (F_{ST}=0.0038, Puebla *et al.* 2014).

Alcolapia diversification and soda lake colonisation

Although it is thought that all *Alcolapia* species currently feed on the same resources of algae and cyanobacteria (Coe 1969; Trewavas 1983), it is likely that there was more extensive trophic niche space available in the deeper palaeolake and *Alcolapia* was not restricted to shallow volcanic springs and lagoon edges. Lake depth (but not lake area) along with energy input (radiation) have been shown to be linked with propensity to diversify in cichlids (Wagner *et al.* 2012). The influence of these factors on diversification has been suggested to be a result of high carrying capacities but also short generation times and increased mutation rates, which are both thought to be factors at play within this system (Wilson *et al.* 2004). Although lake area is not associated with potential to diversify (Wagner *et al.* 2012), for those lakes in which diversification does occur, area predicts number of resultant lineages as adaptive radiation appears to scale with area (Wagner *et al.* 2014). Thus, the restricted lake area in this system may explain why the soda lake radiation contains lower species diversity than seen in radiations from larger lakes. While the current shallow habitat depth for *Alcolapia* (maximum 0.2-1.2 m) negates the benthic-pelagic axis along which freshwater diversification is often seen (e.g., Schliewen *et al.* 2001; Vonlanthen *et al.* 2009; Wagner *et al.* 2012; Praebel *et al.* 2013; Franchini *et al.* 2014), the maximum depth of the palaeolake Orolonga (50-60m; (Roberts *et al.* 1993) was greater than the depth range over which diversification has been recorded in other shallow-water cichlid systems (e.g., Schliewen *et al.* 2001;

Seehausen *et al.* 2008). Thus, adaptation and diversification could have occurred in a deeper, oligosaline lake. In line with this reasoning, the phylogeny is consistent with a scenario in which colonisation of the palaeolake occurred by a freshwater ancestor, with subsequent adaptation to saline/alkaline conditions. Within the deeper water of the palaeolake, divergence would have been possible between terminal mouth morphology (*A. alcalica*) and inferior mouth morphology (*A. ndalalani*/*A. latilabris*) along the major ecological axis of pelagic- or surface-feeding vs. benthic-feeding (Seehausen & Wagner 2014). As the water levels dropped and Lakes Natron/Magadi formed, *A. grahami* would have been geographically isolated from the remaining *Alcolapia* species, while partitioning of ecological niche and divergence of the inferior mouth morphology (thick vs. thin lips) could explain *A. latilabris* and *A. ndalalani* divergence. However, further empirical work would be required to test the ecological and functional relevance of these different trophic morphologies.

As well as being considerably smaller than Lake Natron (covering only ~20% of the area), Lake Magadi also differs in having no perennial inflowing streams, while Lake Natron has two inflowing rivers, Peninj and Ewaso Ngiro, as well as several perennial streams (Olaka *et al.* 2010). This factor not only has implications for hydrochemical variability between the two lakes, but also in terms of niche space available, as inflowing rivers and streams provide longer stretches of continuous open water than the volcanic springs. Furthermore, a previous study recorded differential species distributions along the same stream, with *A. latilabris* found more abundantly in the upper courses (Seegers *et al.* 2001), which could indicate partitioning of habitat use. As such, there may be ecological differences driving genetic differentiation between the Magadi/Natron species as well as the allopatric separation.

Hybridisation within the *Alcolapia* radiation

Species radiations are frequently characterised by interspecific hybridisation after the onset of speciation (Grant & Grant 2008). Here, the f_4 tests provide strong evidence for recent gene flow among all three Lake Natron *Alcolapia* species (Table 3.3). The ML phylogeny also revealed that individuals from a single collecting locality (site 17) did not cluster by species (Figure 3.4). The two samples that exhibited an intermediate form between *A. alcalica* and *A. ndalalani* (causing difficulties with original species designation) grouped with one of the putative parental species (*A. ndalalani*). However, one may have expected the individuals

identified as possible hybrids by phenotype to group separately from parental species. It is possible that, if narrow regions of the genome control traits of colouration and mouth morphology (on which species descriptions are based), hybrid individuals could possess a species-typical phenotype of one parental species while exhibiting a combination of both parental genotypes across the rest of the genome. However, this hypothesis is not tested in the current analysis. Furthermore, while previous studies have shown hybridisation to be an important mechanism in the diversification of other cichlid lineages (Seehausen 2004), additional work would be required to test its role within the soda lake system.

Effect of data quality filtering on phylogenomic signal

As in previous analyses of very young cichlid radiations (Wagner *et al.* 2013), the present analysis demonstrated that inclusion of maximal data from RAD sequencing gave the best phylogenomic resolution, and that removal of additional data in filtering and QC steps (e.g., imposing thresholds for maximum amount of missing data, using higher SNP quality thresholds or higher threshold for individual inclusion) resulted in poorer phylogenomic resolution in ML analysis. This reflects phylogenetic analysis of simulated RAD data, showing that missing data do not appear to negatively impact phylogenetic accuracy, but low levels of informative data may do so (Rubin *et al.* 2012). Thus, the final dataset imposed a minimal threshold for missing data (including all sites with data for >2 individuals), a moderate SNP quality threshold (20) and removed only those individuals with considerably lower sequencing quality than the rest of the dataset (<2.5 million sequenced reads; a total of five individuals). While it may seem intuitively preferable to use maximal stringency in filtering to ensure optimally clean data, several recent analyses have revealed that such stringency may obscure phylogenomic signal in noisy datasets, removing important information in the form of outlier data (Huang & Knowles 2014), and that increased stringency in filtering thresholds may reduce accuracy of downstream analyses (e.g., Hoffman *et al.* 2014)

Phylogenomic inference using next generation sequencing (NGS) datasets

Despite the considerable increase in phylogenetically informative data from NGS, difficulties remain in finding the best way to analyse such large datasets. Of note, the RAD methodology was originally developed for population-level studies (Baird *et*

al. 2008; Hohenlohe *et al.* 2010) and have only recently begun to be used in phylogenomic analysis. The more recent studies that investigate phylogeny using RAD data have typically employed phylogenomic inference using matrices of concatenated sequence data from resultant RAD loci rather than using SNP datasets (e.g., Wagner *et al.* 2013). Phylogenetic algorithms have typically been developed for sequence data and do not handle concatenated SNP data well as it contains no invariant sites (Bertels *et al*. 2014), and may violate assumptions of rate heterogeneity, especially as large datasets often preclude the use of partitioning programs to assign different substitution models to different regions. Using the entire sequence data from RAD sampling (rather than only the variable sites) becomes computationally intensive with increase in taxon number, and it seems that taxa number rather than alignment length is the limiting factor in such analysis. For example, the reduced-taxon dataset (n=25; 26 million bp) took 3 days to complete analysis, whereas the full-taxon dataset (n=92; 28 million bp) would have taken >6 months to complete.

Thus, using only variable sites that are phylogenetically informative seems an intuitive way by which to reduce the dataset to feasible size. However, using a SNP-only dataset reduced the confidence of ML branch partitions, and at least one species-level node (*A. grahami*) was unable to be placed with any confidence, even when using the RAxML ASC model (ascertainment bias correction) suggested for SNP datasets (Appendix Figure 3A.3), while this node was robustly resolved in the analysis of the full sequence data (Figure 3.3). As the reduced-taxon dataset was assembled from the individuals with highest sequence quality in each population, checks were performed to ensure that levels of missing data in the full-taxon SNP-only dataset were not impacting results. However, imposing maximum thresholds of missing data in SNP datasets (maximum 50% and 0% missing data) did not increase bootstrap support or resolve species relationships more robustly (results not shown).

It has been suggested that ascertainment bias correction may not be appropriate for SNP alignments (Pettengill *et al.* 2014), although employing GTRGAMMA with and without the ASC model in the current dataset produced very similar results (Figures 3.3 and 3A.3). Though it should be noted that the results are not directly comparable as these models were run on different datasets given the removal of ambiguous bases from the alignment for use of the ASC model, as the ASC model will not run on alignments containing invariant sites, and sites that contain ambiguous bases are considered invariant if the ambiguous base could be determined to be the same as non-ambiguous bases at that site, and there are no

other polymorphic bases at that site (RAxML 8 manual). While ambiguous bases in Sanger-produced sequences typically indicate lack of confidence in base-calling rather than genuine heterozygosity, those in NGS datasets are usually called only if reaching a minimum coverage in both alternative bases and so are more likely represent a heterozygous locus, and so certain assumptions from existing phylogenetic models may well be unsuited to such SNP datasets.

While some of the difficulties in the use of NGS data for phylogenetic purposes are highlighted here, it is undoubtedly the case that these datasets have already provided resolution to previously unanswered questions, and the continued updating of existing phylogenetic models with SNP-only datasets, along with the development of SNP-specific protocols (e.g., SNAPP program used in the present analysis) will continue to refine and enhance the inferences that can be made from such datasets.

Conclusions

The findings of the present analysis, which demonstrate recent divergence, ongoing gene flow, and low levels of genome-wide divergence alongside narrow peaks of high divergence certainly warrant further investigation in order to elucidate the processes initiating and maintaining speciation in this system. A clear future research goal would be to identify regions of the genome that are responsible for the phenotypic diversification and polymorphism observed in these cichlid fishes, despite the very shallow divergence between species.

References

Aguirre WE, Ellis KE, Kusenda M, Bell MA (2008) Phenotypic variation and sexual dimorphism in anadromous threespine stickleback: implications for postglacial adaptive radiation. *Biological Journal of the Linnean Society*, **95**, 465–478.

Baird NA, Etter PD, Atwood TS *et al.* (2008a) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Barluenga M, Meyer A (2004) The Midas cichlid species complex: incipient sympatric speciation in Nicaraguan cichlid fishes? *Molecular Ecology*, **13**, 2061–2076.

Barluenga M, Stölting KN, Salzburger W, Muschick M, Meyer A (2006) Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*, **439**, 719–723.

Barluenga M, Meyer A (2010) Phylogeography, colonization and population history of the Midas cichlid species complex (*Amphilophus* spp.) in the Nicaraguan crater lakes. *BMC Evolutionary Biology*, **10**, 326.

Bergman AN, Laurent P, Otiang'a-Owiti G *et al.* (2003) Physiological adaptations of the gut in the Lake Magadi tilapia, *Alcolapia grahami*, an alkaline- and saline-adapted teleost fish. *Comparative Biochemistry and Physiology, Part A*, **136**, 701–715.

Bertels F, Silander OK, Pachkov M *et al*. (2014) Automated reconstruction of whole-genome phylogenies from short-sequence reads. *Molecular Biology and Evolution*, **31**, 1077–1088.

Bouckaert RR (2010) DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics*, **26**, 1372–1373.

Bouckaert R, Heled J, Kühnert D *et al.* (2014) BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, **10**, e1003537.

Boulenger GA (1912) Description of a new fish from British East Africa. *Annals and Magazine of Natural History*, **9**, 519–521.

Bryant D (2004) Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, **21**, 255–265.

Bryant D, Bouckaert R, Felsenstein J *et al*. (2012) Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, **29**, 1917–1932.

Burrough SL, Thomas DSG (2009) Geomorphological contributions to palaeolimnology on the African continent. *Geomorphology*, **103**, 285–298.

Cariou M, Duret L, Charlat S (2013) Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecology and Evolution*, **3**, 846–852.

Clayton D, Leung H-T (2007) An R package for analysis of whole-genome association studies. *Human Heredity*, **64**, 45–51.

Coe MJ (1966) The Biology of *Tilapia grahami* Boulenger in Lake Magadi, Kenya. *Acta Tropica*, **23**, 146–198.

Coe MJ (1969) Observations on *Tilapia alcalica* Hilgendorf an endemic cichlid fish from Lake Natron, Tanzania. *Revue de Zoologie et de Botanique Africaines*, **80**, 1–9.

Davey JW, Cezard T, Fuentes-Utrilla P *et al*. (2012). Special features of RAD sequencing data: implications for genotyping. *Molecular Ecology*, **22**, 3151–3164.

De Mita S, Siol M (2012) EggLib: Processing, analysis and simulation tools for population genetics and genomics. *BMC Genetics*, **13**, 27.

DePristo MA, Banks E, Poplin R *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, **43**, 491–498.

Dray S, Dufour AB (2007) The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**, 1–20.

Drummond AJ, Bouckaert R (2014) *Bayesian evolutionary analysis with BEAST 2.* Cambridge University Press.

Earl DA, vonHoldt BM (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: An example from flowering plants (*Pedicularis*: Orobanchaceae). *Systematic Biology*, **62**, 689–706.

Elmer KR, Fan S, Gunter HM *et al.* (2010a) Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes. *Molecular Ecology*, **19**, 197–211.

Elmer KR, Lehtonen TK, Fan S, Meyer A (2012) Crater lake colonization by neotropical cichlid fishes. *Evolution*, **67**, 281–288.

Elmer KR, Lehtonen TK, Kautt AF *et al*. (2010b) Rapid sympatric ecological differentiation of crater lake cichlid fishes within historic times. *BMC Biology*, **8**, 60.

Eschmeyer WN (Ed.) (2015) *Catalog of fishes electronic version. Updated 7 January 2015*. California Academy of Sciences.

Eugster HP (1986) Lake Magadi, Kenya: a model for rift valley hydrochemistry and sedimentation? *Geological Society, London, Special Publications*, **25**, 177–189.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, **1**, 47–50.

Franchini P, Fruciano C, Spreitzer ML *et al.* (2014) Genomic architecture of ecologically divergent body shape in a pair of sympatric crater lake cichlid fishes. *Molecular Ecology*, **23**, 1828–1845.

Gillespie RG (2013) Adaptive radiation: Convergence and non-equilibrium. *Current Biology*, **23**, R71–R74.

Gowell CP, Quinn TP, Taylor EB (2012) Coexistence and origin of trophic ecotypes of pygmy whitefish, *Prosopium coulterii*, in a south-western Alaskan lake. *Journal of Evolutionary Biology*, **25**, 2432–2448.

Grant PR, Grant BR (2008) *How and why species multiply: the radiation of Darwin's finches*. Princeton University Press, Princeton, NJ.

Hilgendorf F (1905) Fische von Deutsche und Englisch Ost-Afrika. Gesammelt von Oxcar Neumann 1893-1895. *Zoologische Jahrbücher*, **22**, 405–420.

Hoffman JI, Simpson F, David P *et al.* (2014) High-throughput sequencing reveals inbreeding depression in a natural population. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 3775–3780.

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.

Huang H, Knowles LL (2014) Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. *Systematic Biology*, **0**, 1–9.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

Hudson AG, Vonlanthen P, Seehausen O (2010) Rapid parallel adaptive radiations from a single hybridogenic ancestral population. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 58–66.

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Johannsson OE, Bergman HL, Wood CM *et al.* (2014) Air breathing in Magadi tilapia *Alcolapia grahami*, under normoxic and hyperoxic conditions, and the association with sunlight and reactive oxygen species. *Journal of Fish Biology*, **84**, 844–863.

Jones BF, Eugster HP, Rettig SL (1977) Hydrochemistry of the Lake Magadi basin, Kenya. *Geochimica et Cosmochimica Acta*, **41**, 53–72.

Jones FC, Chan YF, Schmutz J *et al.* (2012) A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Current biology : CB*, **22**, 83–90.

Kaufman A, Margaritz M, Paul M *et al.* (1990) The ^{36}Cl ages of the brines in the Magadi-Natron basin, East Africa. *Geochimica et Cosmochimica Acta*, **54**, 2827–2833.

Kavembe GD, Machado-Schiaffino G, Meyer A (2013) Pronounced genetic differentiation of small, isolated and fragmented tilapia populations inhabiting the Magadi Soda Lake in Kenya. *Hydrobiologia*, **739**, 55–71.

Keller I, Wagner CE, Greuter L *et al.* (2013) Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology*, **22**, 2848–2863.

Kirchberger PC, Sefc KM, Sturmbauer C, Koblmüller S (2014) Outgroup effects on root position and tree topology in the AFLP phylogeny of a rapidly radiating lineage of cichlid fish. *Molecular Phylogenetics and Evolution*, **70**, 57–62.

Kozak GM, Brennan RS, Berdan EL et al. (2013) Functional and population genomic divergence within and between two species of killifish adapted to different osmotic niches. *Evolution*, **68**, 63–80.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Li H, Handsaker B, Wysoker A *et al.* (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Maina JN (2000) Functional morphology of the gas-gland cells of the air-bladder of *Oreochromis alcalicus grahami* (Teleostei: Cichlidae): an ultrastructural study on a fish adapted to a severe, highly alkaline environment. *Tissue and Cell*, **32**, 117–132.

Mallet J (1995) A species definition for the Modern Synthesis. *Trends in Ecology & Evolution*, **10**, 294–299.

Martin CH, Feinstein LC (2014) Novel trophic niches drive variable progress towards ecological speciation within an adaptive radiation of pupfishes. *Molecular Ecology*, **23**, 1846–1862.

Martin SH, Dasmahapatra KK, Nadeau NJ *et al.* (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, **23**, 1817–1828.

McCormick SD, Farrell AP, Brauner CJ (Eds.) (2013) *Euryhaline Fishes*. Academic Press, Oxford.

McKenna A, Hanna M, Banks E *et al.* (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**, 1297–1303.

Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gateway Computing Environments Workshop 2010..

Nadeau NJ, Whibley A, Jones RT *et al.* (2012) Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **367**, 343–353.

Nagl S, Tichy H, Mayer WE *et al.* (2001) Classification and phylogenetic relationships of African tilapiine fishes inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **20**, 361–374.

Narahara A, Bergman HL, Laurent P *et al.* (1996) Respiratory physiology of the Lake Magadi tilapia (*Oreochromis alcalicus grahami*), a fish adapted to a hot, alkaline, and frequently hypoxic environment. *Physiological Zoology*, **69**, 1114–1136.

Nosil P, Harmon LJ, Seehausen O (2009) Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, **24**, 145–156.

Olaka LA, Odada EO, Trauth MH, Olago DO (2010) The sensitivity of East African rift lakes to climate fluctuations. *Journal of Paleolimnology*, **44**, 629–644.

Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

Pörtner HO, Schulte PM, Wood CM, Schiemer F (2010) Niche dimensions in fishes: an integrative view. *Physiological and Biochemical Zoology*, **83**, 808–826.

Peng Y, Leung HCM, Yiu SM, Chin FYL (2012) IDBA-UD: A *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth.

Bioinformatics, **28**, 1420–1428.

Pettengill JB, Luo Y, Davis S *et al.* (2014) An evaluation of alternative methods for constructing phylogenies from whole genome sequence data: a case study with *Salmonella*. *PeerJ*, **2**, e620.

Praebel K, Knudsen R, Siwertsson A *et al.* (2013) Ecological speciation in postglacial European whitefish: Rapid adaptive radiations into the littoral, pelagic, and profundal lake habitats. *Ecology and Evolution*, **3**, 4970–4986.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Puebla O, Bermingham E, McMillan WO (2014) Genomic atolls of differentiation in coral reef fishes (*Hypoplectrus* spp., *Serranidae*). *Molecular Ecology*, **23**, 5291–5303.

R Core Team (2012) *R: A language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rambaut A, Drummond AJ (2007) *Tracer v1.4*. Available from http://beast.bio.ed.ac.uk/Tracer. Accessed July 2015.

Randall D, Wood CM, Perry SF *et al.* (1989) Urea excretion as a strategy for survival in a fish living in a very alkaline environment. *Nature*, **337**, 165–166.

Reich D, Patterson N, Campbell D *et al.* (2012) Reconstructing Native American population history. *Nature*, **488**, 370–374.

Reusch TBH, Wegner KM, Kalbe M (2001) Rapid genetic divergence in postglacial populations of threespine stickleback (*Gasterosteus aculeatus*): the role of habitat type, drainage and geographical proximity. *Molecular Ecology*, **10**, 2435–2445.

Roberts N, Taieb M, Barker P *et al.* (1993) Timing of the Younger Dryas event in East Africa from lake level changes. *Nature*, **366**, 146–148.

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Rubin BER, Ree RH, Moreau CS (2012) Inferring phylogenies from RAD sequence data. *PLoS ONE*, **7**, e33394.

Sambrook J, Russel DW (2001) Commonly used techniques in molecular cloning. In: *Molecular Cloning, Volume 3*, p. Appendix 8. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schliewen UK, Rassmann K, Markmann M *et al*. (2001) Genetic and ecological divergence of a monophyletic cichlid species pair under fully sympatric conditions in Lake Ejagham, Cameroon. *Molecular Ecology*, **10**, 1471–1488.

Schluter D (2000) *The ecology of adaptive radiation*. Oxford University Press, Oxford.

Schwarzer J, Herder F, Misof B *et al*. (2008) Gene flow at the margin of Lake Matano's adaptive sailfin silverside radiation: Telmatherinidae of River Petea in Sulawesi. *Hydrobiologia*, **615**, 201–213.

Seegers L, Tichy H (1999) The *Oreochromis alcalicus* flock (Teleostei: Cichlidae) from Lake Natron and Magadi, Tanzania and Kenya, with description of two new species. *Ichthyological Explorations of Freshwaters*, **10**, 97–146.

Seegers L, Sonnenberg R, Tichy H (2001) The *Alcolapia* group, a remarkable species flock from lakes Natron, Tanzania, and Magadi, Kenya: a further piece of the puzzle of cichlid evolution? *Journal of Aquariculture and Aquatic Sciences*, **9**, 335–364.

Seegers L, Sonnenberg R, Yamamoto R (1999) Molecular analysis of the *Alcolapia* flock from lakes Natron and Magadi, Tanzania and Kenya (Teleostei: Cichlidae), and implications for their systematics and evolution. *Ichthyological Explorations of Freshwaters*, **10**, 175–199.

Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, **19**, 198–207.

Seehausen O (2006) African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society B: Biological Sciences*, **273**, 1987–1998.

Seehausen O (2013) Conditions when hybridization might predispose populations for adaptive radiation. *Journal of Evolutionary Biology*, **26**, 279–281.

Seehausen O, Wagner CE (2014) Speciation in freshwater fishes. *Annual Review of Ecology, Evolution, and Systematics*, **45**, 621–651.

Seehausen O, Terai Y, Magalhaes IS *et al.* (2008) Speciation through sensory drive in cichlid fish. *Nature*, **455**, 620–626.

Sültmann H, Mayer WE, Figueroa F, Tichy H, Klein J (1995). Phylogenetic analaysis of cichlid fishes using nuclear DNA markers. *Molecular Biology and Evolution*, **12**, 1033–1047.

Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Tebbs EJ, Remedios JJ, Avery ST, Harper DM (2013) Remote sensing the hydrological variability of Tanzania's Lake Natron, a vital Lesser Flamingo breeding site under threat. *Ecohydrology & Hydrobiology*, **13**, 148–158.

Templeton AR (1989) The meaning of species and speciation: A genetic perspective. In: *Speciation and its consequences* (Eds Otte D, Endler JA), pp.

3–27. Sinauer Associates, Sunderland, MA.

Tichy H, Seegers L (1999) The *Oreochromis alcalicus* flock (Teleostei Cichlidae) from lakes Natron and Magadi, Tanzania and Kenya: A model for the evolution of new species flocks in historical times. *Ichthyological Explorations of Freshwaters*, **10**, 147–174.

Trewavas E (1983) *Tilapiine Fishes of the genera Sarotherodon, Oreochromis and Danakilia*. British Museum (Natural History), London.

Vanden Bossche J-P, Bernacsek GM (1990) *Source book for the inland fishery resources of Africa 1. CIFA Technical Paper. No. 18.1.* Food and Agriculture Organisation, Rome.

Vareschi, E. (1979). The ecology of Lake Nakuru (Kenya). II. Biomass and spatial distribution of fish (*Tilapia grahami* Boulenger = *Sarotherodon alcalicum grahami* Boulenger). *Oecologia*, **37**, 321–335.

Vincenty T (1975) Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey Review*, **22**, 1–7.

Vonlanthen P, Bittner D, Hudson AG *et al.* (2012) Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature*, **482**, 357–362.

Vonlanthen P, Roy D, Hudson AG *et al.* (2009) Divergence along a steep ecological gradient in lake whitefish (*Coregonus* sp.). *Journal of Evolutionary Biology*, **22**, 498–514.

Wagner CE, Harmon LJ, Seehausen O (2012) Ecological opportunity and sexual selection together predict adaptive radiation. *Nature*, **487**, 366–369.

Wagner CE, Harmon LJ, Seehausen O (2014) Cichlid species-area relationships are shaped by adaptive radiations that scale with area. *Ecology Letters*, **17**, 583–592.

Wagner CE, Keller I, Wittwer S *et al.* (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**, 787–798.

Walsh PJ, Grosell M, Goss GG *et al.* (2001) Physiological and molecular characterization of urea transport by the gills of the Lake Magadi tilapia (*Alcolapia grahami*). *Journal of Experimental Biology*, **204**, 509–520.

Williamson D, Taieb M, Damnati B *et al*. (1993) Equatorial extension of the Younger Dryas event rock magnetic evidence from Lake Magadi (Kenya). *Global and Planetary Change*, **7**, 235–242.

Wilson PJ, Wood CM, Maina JN, White BN (2000). Genetic structure of Lake Magadi tilapia populations. *Journal of Fish Biology*, **56**, 590–603.

Wilson PJ, Wood CM, Walsh PJ *et al.* (2004) Discordance between genetic structure and morphological, ecological, and physiological adaptation in Lake Magadi tilapia. *Physiological and Biochemical Zoology*, **77**, 537–555.

Wood CM, Bergman HL, Laurent P *et al.* (1994) Urea production, acid-base regulation and their interactions in the Lake Magadi tilapia, a unique teleost adapted to a highly alkaline environment. *Journal of Experimental Biology*, **189**, 13–36.

Zaccara S, Crosa G, Vanetti I *et al.* (2014) Genetic and morphological analyses indicate high population mixing in the endangered cichlid *Alcolapia* flock of East Africa. *Conservation Genetics*, **15**, 429–440.

Appendix Three

Appendix 3A.1 Phenol/Chloroform extraction protocol, modified from (Sambrook & Russel 2001)

1. Place tissue sample in 1.5mL Eppendorff tube with 500µl extraction buffer with SDS (recipe below) and 15µL Proteinase K
2. Incubate at 37˚C for 6 hours or 55˚C for 4 hours. [RAD: 55˚C for 1 hour]
3. Vortex and spin down
4. Add 4µL RNAse A and incubate for 2 minutes
5. Mix by overhead inversion for 10 minutes
6. Add 75µL NaCl 5M and mix by inversion for 10 mins
7. Add 650µL Phenol:Chloroform:Isoamyl Alcohol (25:24:1) – in fume cupboard
8. Mix by overhead inversion for 10 mins
9. Centrifuge: 15 mins at 10,000 RPM
10. Remove supernatant phase to new tube – in fume cupboard
11. Add 1000µL ice cold 100% ethanol (stored in freezer)
12. Put samples in freezer -20 ˚C for >1 hour
13. Centrifuge: 15 mins at 15,000 RPM at 4 ˚C
14. Pour off ethanol (being careful not to dislodge pellet at bottom of tube)
15. Add 1500µL ice cold 70% ethanol (from freezer)
16. Centrifuge: 10 mins at 13,000 RPM 4 ˚C
17. Pour off ethanol (taking care not to dislodge pellet)
18. Dry samples on hot blocks at 37 ˚C for ~15 mins
19. Resuspend pellet in 50-75µL distilled H20 or DNA storage buffer (TE)

Extraction buffer (100 mL)

- 10mL Tris-HCl 1M (pH=8)
- 2mL EDTA 0.5M
- 2mL NaCl 5M
- 0.5mL SDS 20%
- 85.5mL dH_20

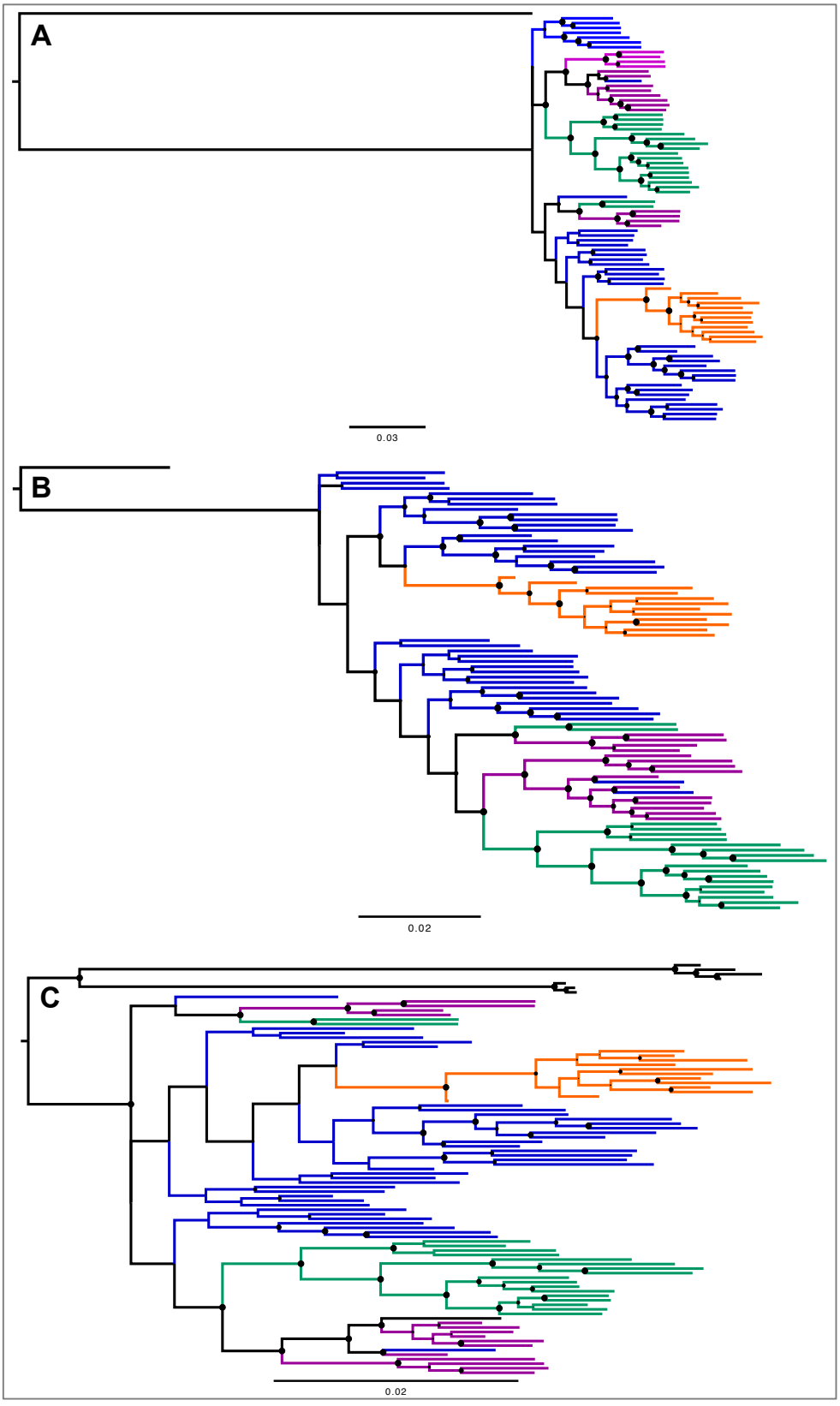


Figure 3A.1. ML analysis using different genotype quality thresholds.

A) GQ 10; B) GQ 20; C) GQ 30. Branches to outgroups (black) have been truncated for clarity. Branches are coloured by species and bootstrap values (100 BS) are indicated by black circles, scaled by support value.

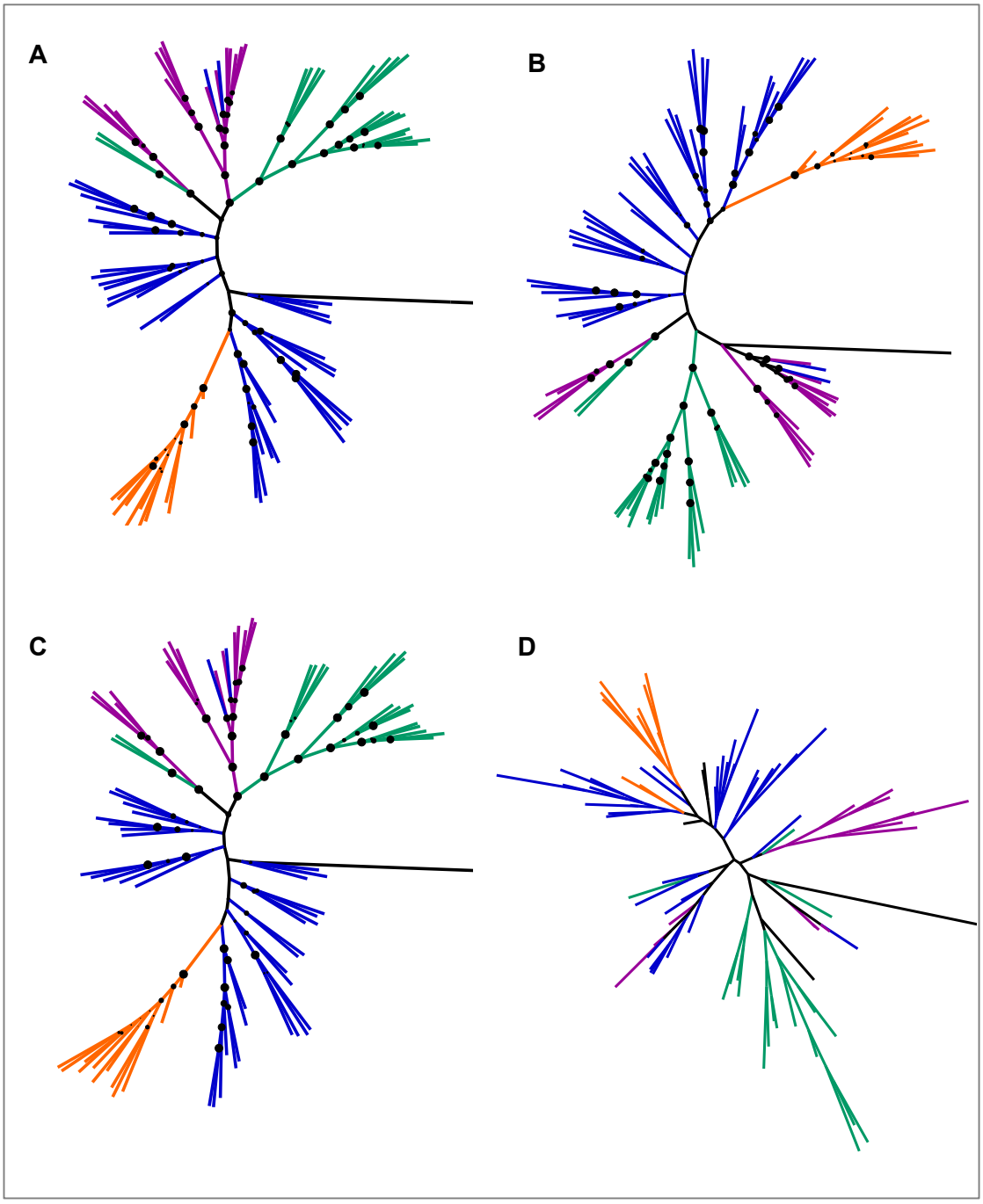


Figure 3A.2. ML analysis using different levels of missing data thresholds.

A) No threshold; B) 75% threshold; C) 25% threshold; D) 0% threshold (no missing data tolerated). Branches to outgroups (black) have been truncated for clarity. Branches are coloured by species and bootstrap values (100 BS) are indicated by black circles, scaled by support value.

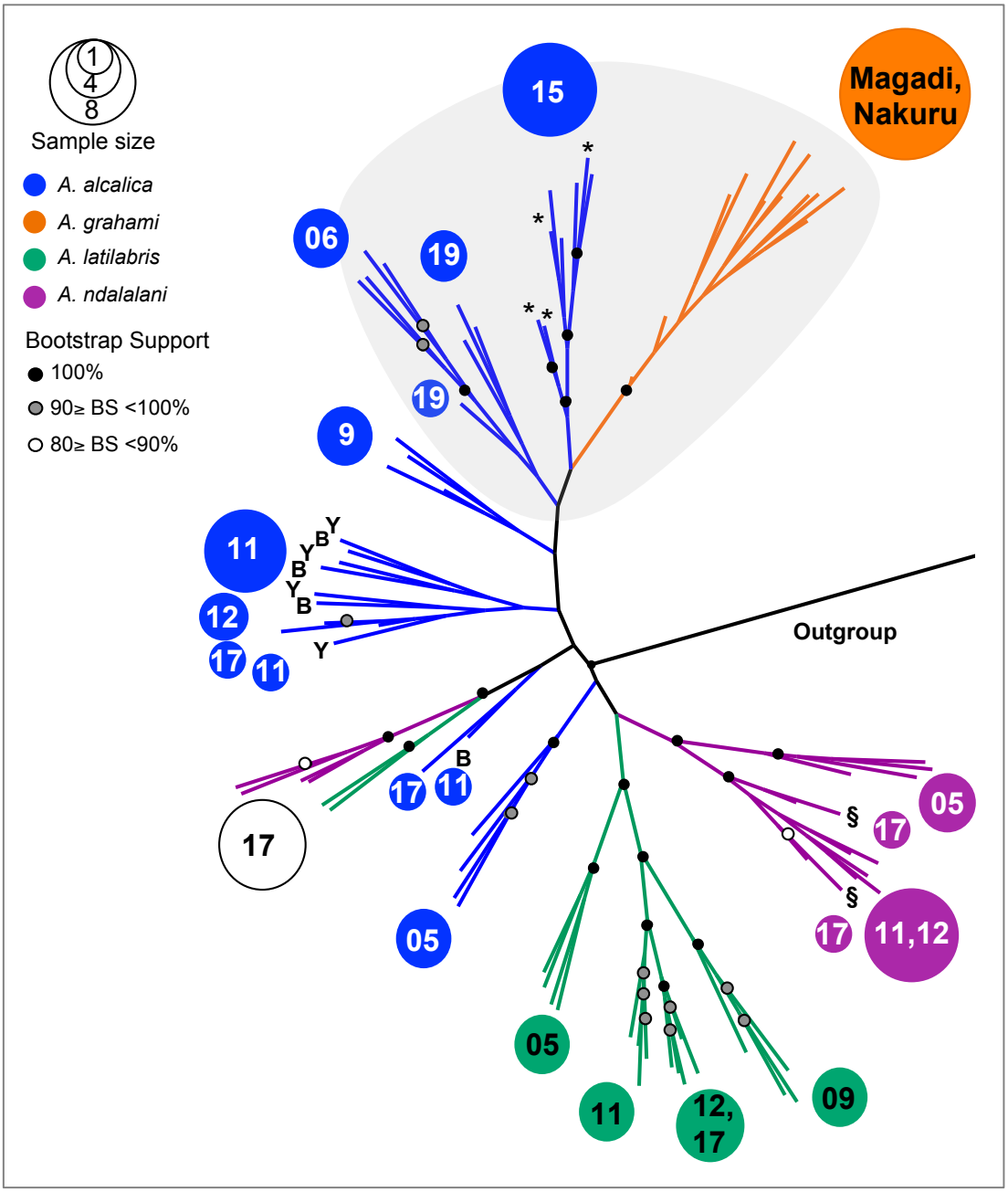


Figure 3A.3. ML phylogeny for variable sites using ASC model.

Full taxon dataset (n=92) alignment of variable sites only with ambiguous bases removed (dataset E; 363,983 SNPs) and ascertainment bias correction (ASC model) applied within RAxML. The phylogeny exhibits similar topology as the phylogeny produced when the ML analysis is conducted without the ASC model (Figure 3.3C in the main chapter text).

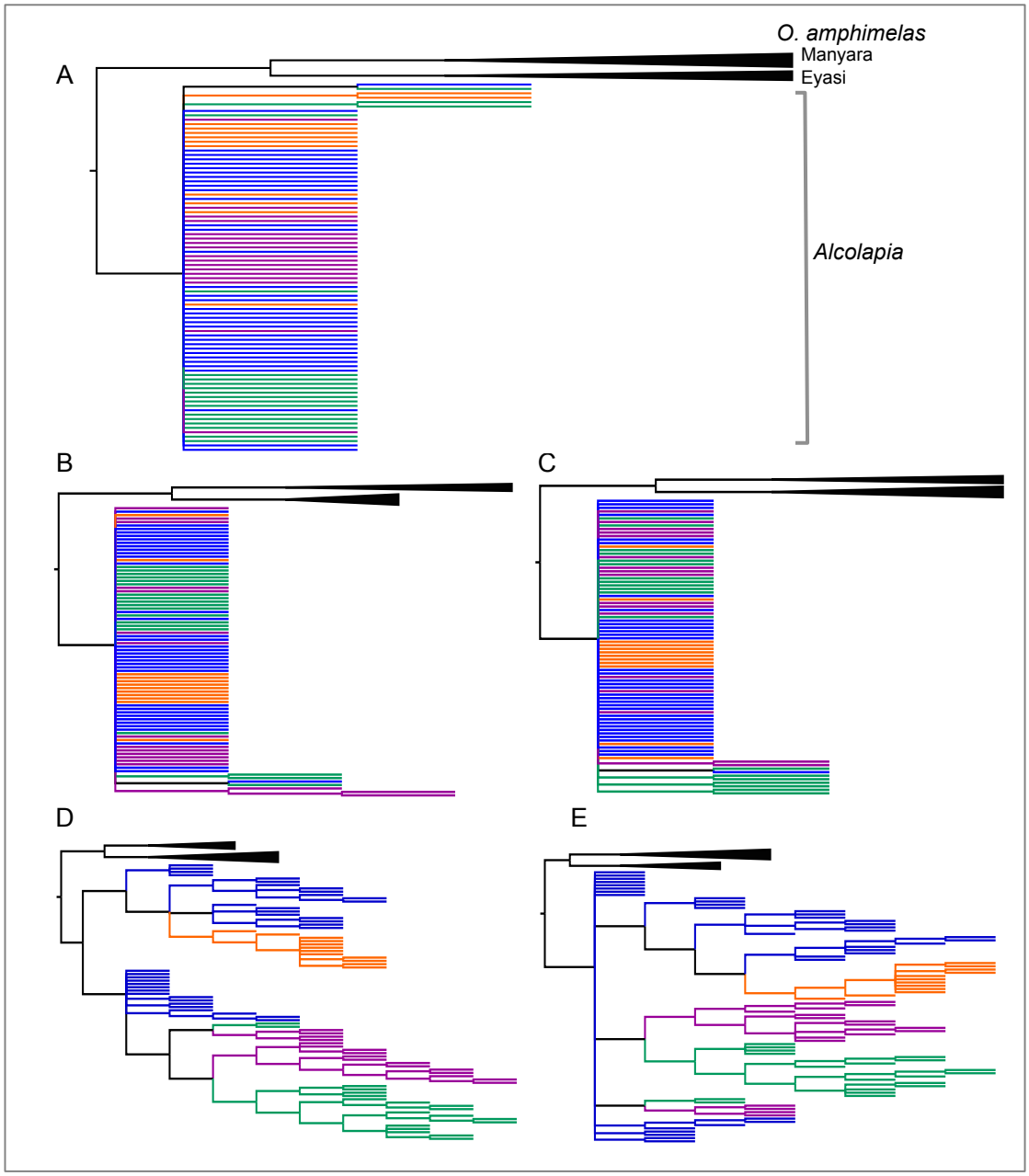


Figure 3A.4. Majority consensus (50%) ML phylogenies generated from additional RAD datasets.

For all analyses, trees are rooted using *O. amphimelas* as outgroup. A) *de novo* assembly (unmapped reads) full dataset (total alignment of 436,839 bp, including invariant sites); B, C) *de novo* assembly variable sites only (SNPs) dataset, B) all SNPs, without ASC correction (5,832 SNPS); C) ambiguous bases removed and employing ASC correction (1,898 SNPs); D, E) combined dataset of SNPs from mapped and *de novo*-assembled reads; D) all SNPs, without ASC correction (550,748 SNPS); C) ambiguous bases removed and employing ASC correction (182,909 SNPs).

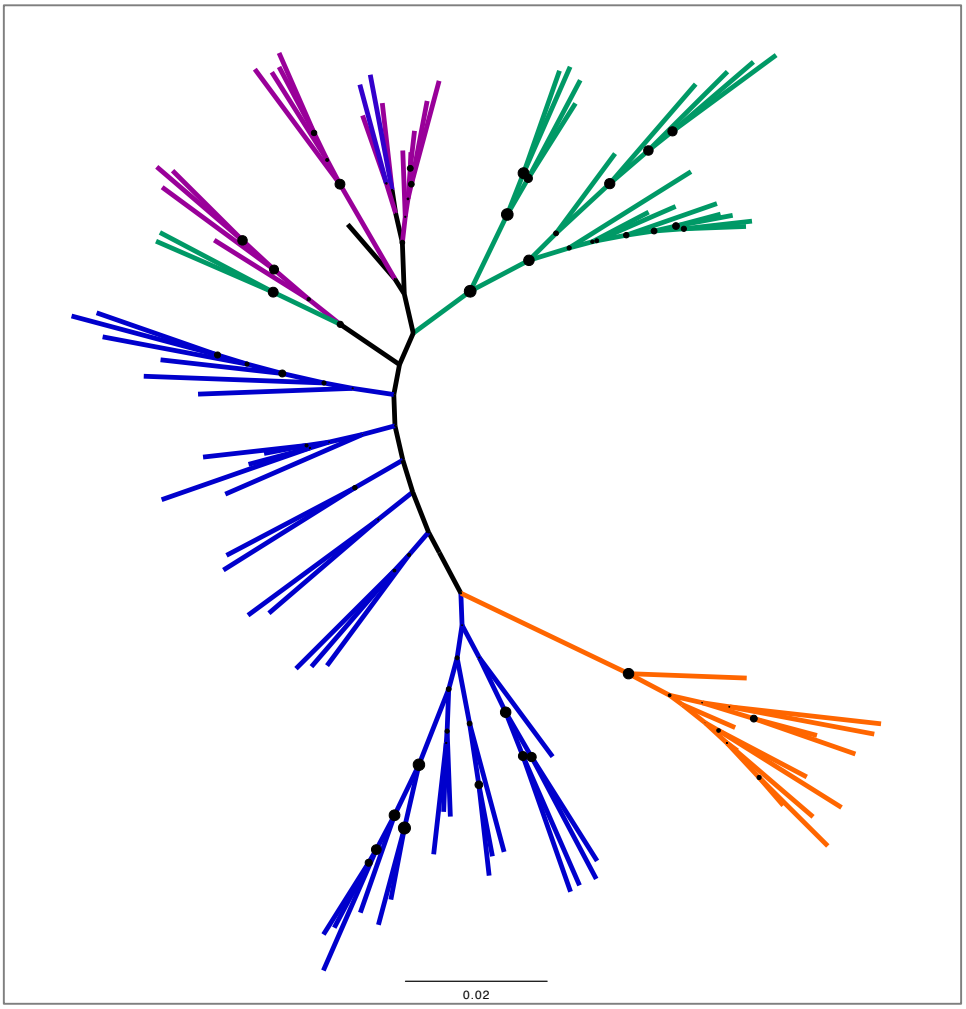


Figure 3A.5. ML analysis of *Alcolapia* excluding outgroup.

RAxML analysis of the dataset excluding outgroups (*O. niltoticus* or *O. amphimelas*) exhibited the same topology as analyses with the outgroup (Figure 3.3; 3A.3), suggesting the outgroup branch placement does not influence ingroup relationships.

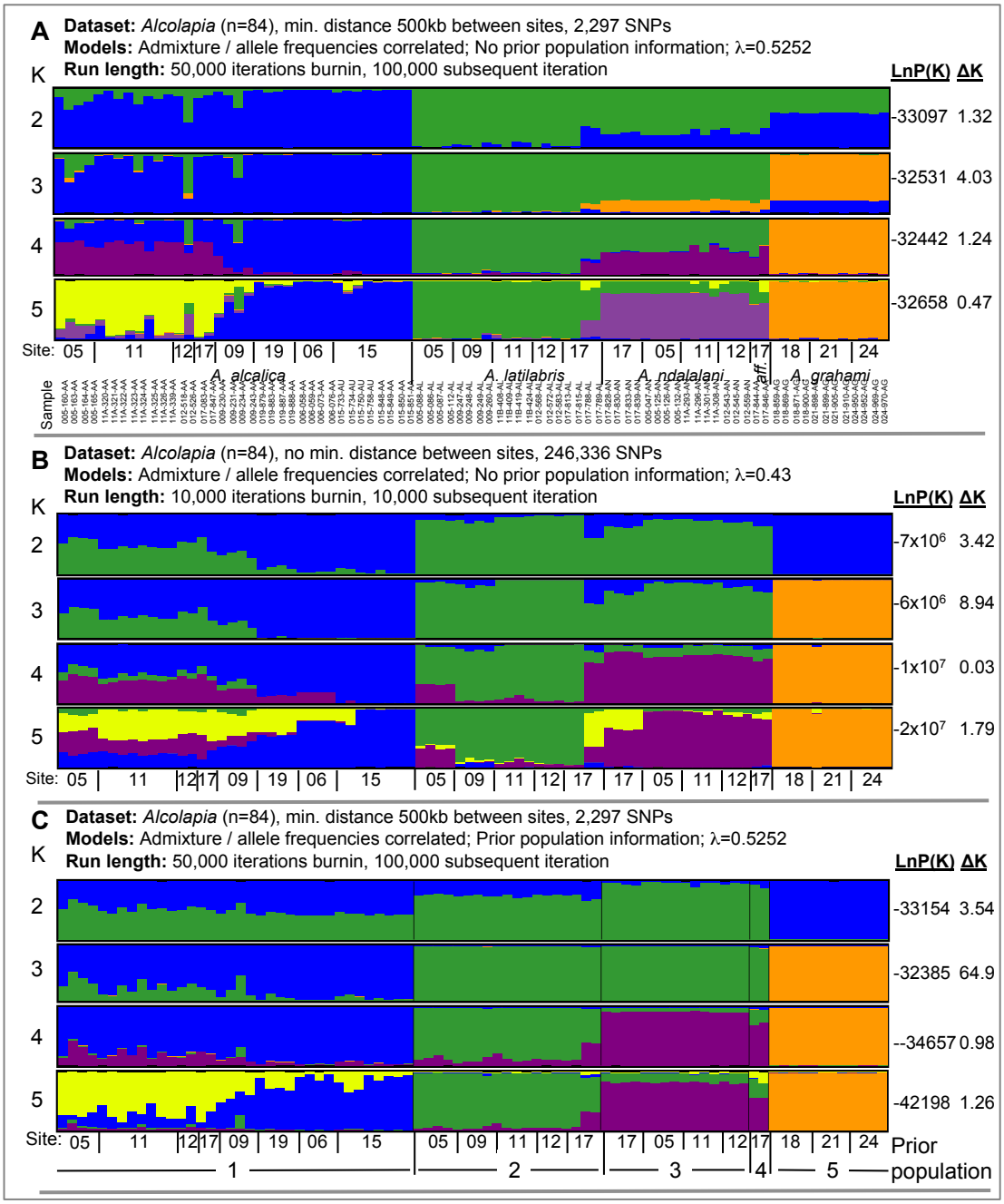


Figure 3A.6. Visualisation of K=2-5 for *Alcolapia* STRUCTURE analysis.

Cluster membership for admixture/correlated allele frequency models. A) Reduced dataset accounting for LD; B) Full dataset, no min. distance between SNPs; C) Population information included as a model prior. Samples are ordered as in panel A. LnP(K): Estimated Ln Prob of Data. ΔK: second order rate of change of the likelihood function with respect to K, the modal value of which indicates the best estimate of K.

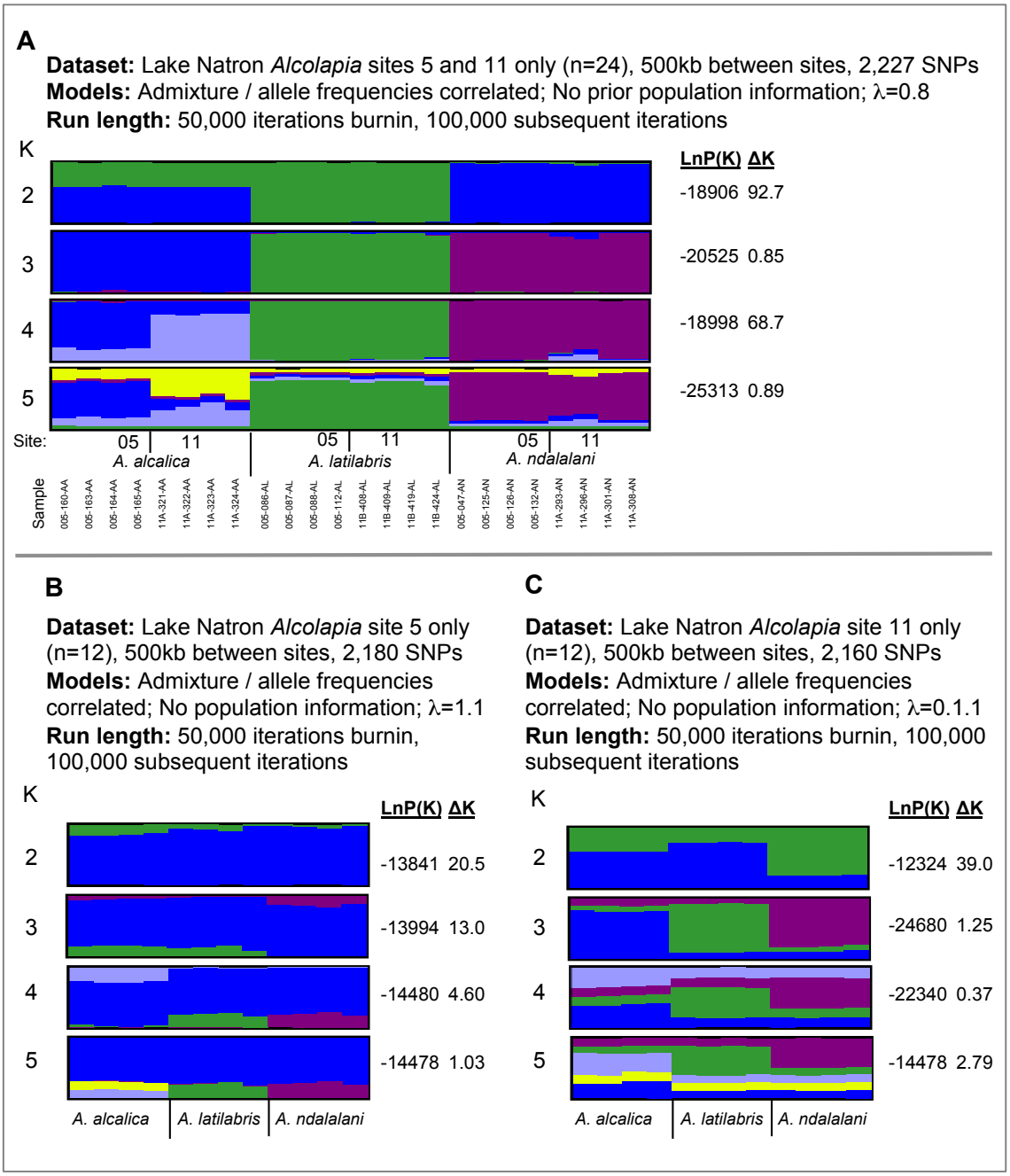


Figure 3A.7. Visualisation of K=2-5 for *Alcolapia* STRUCTURE analysis.

Cluster membership by individual for admixture / correlated allele frequency models for sites in which all three Lake Natron species occur sympatrically: A) Sites 005 and 011 Individuals combined; B) Site 005 only; C) Site 011 only.

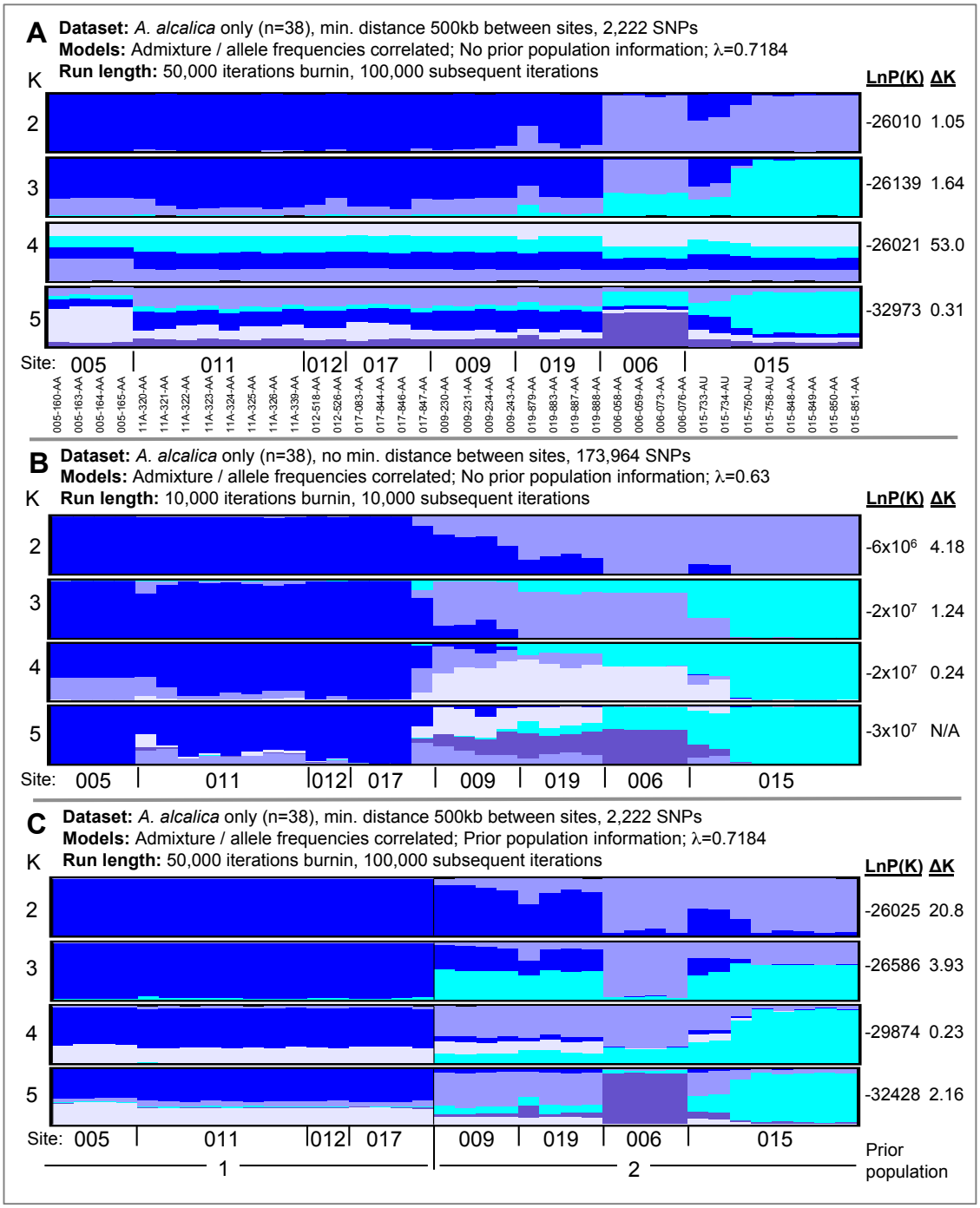


Figure 3A.8. Visualisation of K=2-5 for *A. alcalica* STRUCTURE analysis. Cluster membership for admixture/correlated allele frequency models. A) Reduced dataset accounting for LD; B) Full dataset with no min. distance between SNPs; C) Providing population information as a model prior. Samples ordered as indicated in panel A.

Table 3A.1. Collection co-ordinates and sequencing statistics per sample.

Samples in grey were excluded from downstream analyses due to poor quality and low total calls in the filtered dataset. OA: *O. amphimelas*; AA: *A. alcalica*; AB: *A. alcalica* blue morph; AY: *A. alcalica* yellow morph; AU: *A. alcalica* upturned-mouth morph; AL: *A. latilabris*; AN: *A. ndalalani*; AG: *A. grahami*

	Sample ID	Accession reference ID	Sampling GPS coordinates	Sequenced reads (n)	Duplicates (n)	Duplicates (%)	Mapped (%)	Paired (%)	Total calls (n)	QC-filtered calls (n)	QC-filtered calls (%)	Mean mapping rate
O. amphimelas	002-001-OA	ERS652626	S -3.423333 E 35.851806	1357822	365705	26.93	89.73	81.15	19225625	N/A	N/A	
	002-003-OA	ERS652628	S -3.423333 E 35.851806	3760790	1122956	29.86	89.67	80.88	27501927	12351820	44.91	
	002-004-OA	ERS652623	S -3.423333 E 35.851806	8688110	3077792	35.43	89.73	80.78	32790200	18166870	55.40	*O. amphimelas*
	002-005-OA	ERS652624	S -3.423333 E 35.851806	5593242	1797032	32.13	89.72	80.95	28857245	15351777	53.20	
	004-029-OA	ERS652650	S -3.425306 E 35.343723	8499968	2816835	33.14	89.85	80.55	34897625	18049330	51.72	
	004-030-OA	ERS652670	S -3.425306 E 35.343723	4614448	1828295	39.62	89.85	81.38	25410823	12624979	49.68	89.85
	004-032-OA	ERS652669	S -3.425306 E 35.343723	6429908	2044339	31.79	90.12	81.22	29516848	16229295	54.98	
	004-035-OA	ERS652671	S -3.425306 E 35.343723	6050738	1933421	31.95	90.16	81.46	30160099	16031277	53.15	
A. alcalica	005-160-AA	ERS652662	S -2.597583 E 35.918417	7962738	3157400	39.65	89.47	78.23	28647369	16339639	57.04	
	005-163-AA	ERS652663	S -2.597583 E 35.918417	12639506	5344313	42.28	89.40	78.24	31790644	19339959	60.84	
	005-164-AA	ERS652709	S -2.597583 E 35.918417	7767920	3151245	40.57	89.54	78.38	28511977	16007180	56.14	
	005-165-AA	ERS652665	S -2.597583 E 35.918417	6919164	2706673	39.12	89.42	78.62	28611615	15460588	54.04	
	006-058-AA	ERS652659	S -2.430397 E 35.895392	9480270	3404504	35.91	89.87	79.22	29482851	18084493	61.34	
	006-059-AA	ERS652658	S -2.430397 E 35.895392	8037504	2801660	34.86	89.62	78.93	29306108	17263844	58.91	
	006-073-AA	ERS652686	S -2.430397 E 35.895392	7386516	2595984	35.14	89.77	78.90	29135230	16782296	57.60	
	006-076-AA	ERS652684	S -2.430397 E 35.895392	8719750	3142721	36.04	89.75	78.87	29901941	17808281	59.56	
	009-230-AA	ERS652618	S -2.471278 E 35.887889	1627594	572245	35.16	89.42	78.94	19084112	4587454	24.04	
	009-231-AA	ERS652619	S -2.471278 E 35.887889	4312168	1629393	37.79	89.45	78.84	25365205	11775955	46.43	*A. alcalica*
	009-234-AA	ERS652617	S -2.471278 E 35.887889	10127960	4226515	41.73	89.31	78.18	31697682	17881771	56.41	
	009-243-AA	ERS652625	S -2.471278 E 35.887889	9616338	3959289	41.17	89.20	78.21	29774635	17480225	58.71	
	11A-320-AB	ERS652657	S -2.591000 E 36.009445	2475984	970742	39.21	89.87	79.55	21348756	7033004	32.94	88.79
	11A-321-AB	ERS652656	S -2.591000 E 36.009445	6274036	2630185	41.92	89.74	79.41	26350715	14488129	54.98	
	11A-322-AB	ERS652655	S -2.591000 E 36.009445	9687160	4304246	44.43	89.77	79.23	28644508	17112417	59.74	
	11A-323-AY	ERS652654	S -2.591000 E 36.009445	11694544	5231031	44.73	89.82	79.34	29542463	18514362	62.67	
	11A-324-AY	ERS652653	S -2.591000 E 36.009445	8685264	3844753	44.27	89.82	79.44	27323911	16633029	60.87	
	11A-325-AY	ERS652652	S -2.591000 E 36.009445	9991060	4298967	43.03	89.57	78.88	27581617	17649780	63.99	
	11A-326-AY	ERS652651	S -2.591000 E 36.009445	10447384	4549111	43.54	89.61	79.07	28316443	17775105	62.77	
	11A-339-AB	ERS652672	S -2.591000 E 36.009445	10093142	4534603	44.93	89.74	79.02	27169435	17340274	63.82	
	012-517-AA	ERS652673	S -2.618972 E 35.999806	2150132	1409805	65.57	87.53	75.95	14321762	1203496	8.40	
	012-518-AA	ERS652674	S -2.618972 E 35.999806	9564398	6392842	66.84	87.09	75.22	26391555	8397909	31.82	
	012-521-AA	ERS652638	S -2.618972 E 35.999806	1193110	772867	64.78	86.67	75.71	9998081	313244	3.13	
	012-526-AA	ERS652641	S -2.618972 E 35.999806	4300436	2864522	66.61	87.42	75.84	19390154	3343132	17.24	

	Sample ID	Accession reference ID	Sampling GPS coordinates	Sequenced reads (n)	Duplicates (n)	Duplicates (%)	Mapped (%)	Paired (%)	Total calls (n)	QC-filtered calls (n)	QC-filtered calls (%)	Mean mapping rate
A. alcalica	015-733-AU	ERS652635	S -2.433361 E 36.10175	6589102	4432160	67.27	87.55	76.23	23136938	5559416	24.03	
	015-734-AU	ERS652700	S -2.433361 E 36.10175	24249950	15684157	64.68	84.99	72.88	34712858	19067432	54.93	
	015-750-AU	ERS652664	S -2.433361 E 36.10175	5625700	3608042	64.13	86.92	75.56	22447628	5440646	24.24	
	015-758-AU	ERS652642	S -2.433361 E 36.10175	12987418	7517598	57.88	89.26	78.61	28077266	16124699	57.43	
	015-848-AA	ERS652630	S -2.433361 E 36.10175	8412214	4502353	53.52	88.83	78.30	26656841	16556078	62.11	
	015-849-AA	ERS652631	S -2.433361 E 36.10175	7886154	4430135	56.18	88.95	78.50	25690733	16016409	62.34	*A. alcalica*
	015-850-AA	ERS652644	S -2.433361 E 36.10175	6311730	3354039	53.14	88.88	78.46	24020411	14792394	61.58	
	015-851-AA	ERS652643	S -2.433361 E 36.10175	9803922	5116278	52.19	88.62	77.64	26692651	15371665	57.59	
	017-083-AA	ERS652697	S -2.456278 E 36.087806	7439702	2553925	34.33	89.64	78.43	29072971	13810736	47.50	88.79
	017-844-AA	ERS652634	S -2.456278 E 36.087806	12395920	7039379	56.79	89.05	78.41	28085487	12560157	44.72	
	017-846-AA	ERS652632	S -2.456278 E 36.087806	10438592	5865487	56.19	89.20	78.48	27009981	11431905	42.32	
	017-847-AA	ERS652633	S -2.456278 E 36.087806	11246308	6332246	56.31	89.10	78.40	27327923	15068717	55.14	
	019-879-AA	ERS652680	S -2.145833 E 36.05575	27008098	17754321	65.74	87.20	76.41	31615471	18962739	59.98	
	019-883-AA	ERS652701	S -2.145833 E 36.05575	11096528	7197186	64.86	87.86	76.91	24736152	12254593	49.54	
	019-887-AA	ERS652702	S -2.145833 E 36.05575	4305954	2473368	57.44	88.30	77.71	21038958	7229892	34.36	
	019-888-AA	ERS652703	S -2.145833 E 36.05575	8926152	5281875	59.17	86.54	75.89	25180896	12213631	48.50	
A. latilabris	005-086-AL	ERS652705	S -2.597583 E 35.918417	5299794	1723022	32.51	89.72	78.99	26629605	14388660	54.03	
	005-087-AL	ERS652704	S -2.597583 E 35.918417	7195264	2417648	33.60	89.59	78.77	28951967	16735599	57.80	
	005-088-AL	ERS652706	S -2.597583 E 35.918417	9458188	3352199	35.44	89.58	78.76	29921568	18547005	61.99	
	005-112-AL	ERS652647	S -2.597583 E 35.918417	4150534	1552605	37.41	89.32	78.49	24748665	11398690	46.06	
	009-247-AL	ERS652627	S -2.471278 E 35.887889	11118486	4640657	41.74	89.54	78.28	31325719	18615031	59.42	
	009-248-AL	ERS652621	S -2.471278 E 35.887889	12354302	5261344	42.59	89.57	78.42	31524470	19376823	61.47	
	009-249-AL	ERS652620	S -2.471278 E 35.887889	8642950	3539857	40.96	89.59	78.48	29261691	16895633	57.74	
	009-260-AL	ERS652694	S -2.471278 E 35.887889	8318970	3321266	39.92	89.51	78.68	29051678	16707856	57.51	
	11B-408-AL	ERS652668	S -2.591667 E 36.012056	4279346	1763096	41.20	89.84	79.39	23323430	11318570	48.53	*A. latilabris*
	11B-409-AL	ERS652667	S -2.591667 E 36.012056	8910514	3846933	43.17	89.76	78.92	26603729	16650052	62.59	
	11B-419-AL	ERS652679	S -2.591667 E 36.012056	4204078	1722750	40.98	89.88	79.63	23804721	11273496	47.36	
	11B-424-AL	ERS652689	S -2.591667 E 36.012056	4836192	3148127	65.10	86.67	75.34	21110867	4204765	19.92	88.74
	012-567-AL	ERS652681	S -2.618972 E 35.999806	1479758	947123	64.01	85.34	74.07	11342001	487321	4.30	
	012-568-AL	ERS652685	S -2.618972 E 35.999806	11460052	7544751	65.84	87.39	75.79	28122043	11200347	39.83	
	012-572-AL	ERS652695	S -2.618972 E 35.999806	5437930	3517713	64.69	86.63	75.16	22682601	4851003	21.39	
	012-583-AL	ERS652649	S -2.618972 E 35.999806	6071594	4054002	66.77	87.14	75.69	22961856	4980778	21.69	
	017-788-AL	ERS652640	S -2.456278 E 36.087806	20111250	11830129	58.82	89.05	78.20	30313472	19209844	63.37	
	017-789-AL	ERS652639	S -2.456278 E 36.087806	17481210	10205842	58.38	89.15	78.40	29486809	18331005	62.17	
	017-813-AL	ERS652688	S -2.456278 E 36.087806	11007076	6154940	55.92	88.95	78.27	28032672	15308624	54.61	
	017-815-AL	ERS652687	S -2.456278 E 36.087806	6506730	3397642	52.22	88.59	77.86	25160380	11882911	47.23	

	Sample ID	Accession reference ID	Sampling GPS coordinates	Sequenced reads (n)	Duplicates (n)	Duplicates (%)	Mapped (%)	Paired (%)	Total calls (n)	QC-filtered calls (n)	QC-filtered calls (%)	Mean mapping rate
A. ndalalani	005-047-AN	ERS652660	S -2.597583 E 35.918417	3370746	1030405	30.57	89.68	78.90	24549929	10930759	44.52	
	005-125-AN	ERS652710	S -2.597583 E 35.918417	9426726	3863169	40.98	89.21	77.95	29691805	17517360	59.00	
	005-126-AN	ERS652711	S -2.597583 E 35.918417	6624882	2595957	39.18	89.37	78.15	27799292	14926950	53.70	
	005-132-AN	ERS652629	S -2.597583 E 35.918417	7227458	2866770	39.66	89.26	78.14	28531493	15518207	54.39	
	11A-293-AN	ERS652699	S -2.591000 E 36.009445	3920796	1574459	40.16	89.29	78.86	22827759	10935424	47.90	
	11A-296-AN	ERS652698	S -2.591000 E 36.009445	6773356	2775470	40.98	89.67	79.26	26283459	15225966	57.93	
	11A-301-AN	ERS652645	S -2.591000 E 36.009445	5475056	2199871	40.18	89.39	78.79	24466137	13467064	55.04	*A. ndalalani*
	11A-308-AN	ERS652616	S -2.591000 E 36.009445	1739654	676957	38.91	89.92	79.92	18722285	4795225	25.61	
	012-543-AN	ERS652707	S -2.618972 E 35.999806	5338238	3589974	67.25	87.37	75.91	20923070	4272347	20.42	88.75
	012-545-AN	ERS652708	S -2.618972 E 35.999806	7347420	4936444	67.19	87.44	75.93	23581234	6377516	27.04	
	012-559-AN	ERS652622	S -2.618972 E 35.999806	5496640	3561925	64.80	87.30	75.83	22386833	5164868	23.07	
	012-562-AN	ERS652682	S -2.618972 E 35.999806	2752934	1745377	63.40	86.95	75.61	17204226	2166627	12.59	
	017-828-AN	ERS652661	S -2.456278 E 36.087806	15256986	8462298	55.47	88.72	77.68	29291995	17849726	60.94	
	017-830-AN	ERS652677	S -2.456278 E 36.087806	3343728	1756241	52.52	88.95	78.57	20432579	6679634	32.69	
	017-833-AN	ERS652675	S -2.456278 E 36.087806	11002348	5598092	50.88	88.45	77.71	28055294	16269760	57.99	
	017-839-AN	ERS652678	S -2.456278 E 36.087806	16709740	9782809	58.55	89.09	78.19	29204312	17900731	61.29	
A. grahami	018-859-AG	ERS652693	S -2.001139 E 36.231972	11332104	7316137	64.56	82.42	72.31	23007864	10787850	46.89	
	018-869-AG	ERS652696	S -2.001139 E 36.231972	4807842	3357909	69.84	88.68	77.91	18492529	4395507	23.77	
	018-871-AG	ERS652683	S -2.001139 E 36.231972	38009624	25200827	66.30	87.96	77.04	33738637	21316544	63.18	
	018-900-AG	ERS652636	S -2.001139 E 36.231972	7910266	5012316	63.36	88.45	77.57	23267126	10405216	44.72	
	021-898-AG	ERS652692	S -1.844444 E 36.22425	2163910	1241469	57.37	87.69	77.22	16588390	2926140	17.64	*A. grahami*
	021-899-AG	ERS652691	S -1.844444 E 36.22425	37937308	27520245	72.54	88.14	76.64	29148312	18680203	64.09	
	021-905-AG	ERS652637	S -1.844444 E 36.22425	22813336	14797603	64.86	87.29	76.42	28182867	17694293	62.78	
	021-910-AG	ERS652646	S -1.844444 E 36.22425	4502606	2915439	64.75	88.26	77.65	19456528	5554263	28.55	86.70
	024-950-AG*	ERS652676	S -0.396028 E 36.107583	8081388	4910222	60.76	87.56	76.93	24052205	11267314	46.85	
	024-952-AG*	ERS652690	S -0.396028 E 36.107583	3456626	1932130	55.90	87.36	76.91	20057343	5936464	29.60	
	024-969-AG*	ERS652648	S -0.396028 E 36.107583	4365002	2897370	66.38	85.52	74.94	18065966	3914662	21.67	
	024-970-AG*	ERS652666	S -0.396028 E 36.107583	9962378	4713690	47.31	84.30	74.34	26614209	15712608	59.04	
	TOTAL:			836345864	439397761	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	MEAN:			8711936	4577060	49.75	88.64	77.94	25838916	12850898	47.10	N/A

*indicates sample collected from introduced population at Lake Nakuru. Voucher specimens (unaccessioned) are held in the Day lab at UCL.

Chapter four

Patterns of genomic differentiation in the *Alcolapia* species flock

Abstract

Understanding how speciation occurs in the face of ongoing gene flow has received considerable research focus in recent decades. The concept of genomic islands of speciation, where narrow regions of the genome exhibit high levels of differentiation and are resistant to introgression, is particularly pertinent in incipient species where levels of gene flow are thought to be considerable. Here, the increased power of densely sampled SNPs allows identification of genomic peaks of differentiation (F_{ST} outliers) between *Alcolapia* species. While evidence of ongoing gene flow and interspecies hybridisation in certain populations suggests that *Alcolapia* species are incompletely reproductively isolated (chapter three), the identification of outlier SNPs under diversifying selection indicates that the radiation is undergoing adaptive divergence.

Introduction

The build up of variation within natural populations remains a key research focus in evolutionary biology. The processes contributing to speciation and mechanisms that allow divergence of lineages in the face of gene flow have been of increasing interest in the last few decades (discussed in Faria *et al.* 2014). The recognition that most speciation models involve some level of gene flow throughout the speciation process (whether sympatric/parapatric or secondary contact) has increased focus on the ecological and adaptive mechanisms that may drive speciation-with-gene-flow processes (Smadja & Butlin 2011; Marie Curie Speciation Network, 2012; Faria *et al.* 2014). As well as identifying the components contributing to reproductive isolation despite gene flow, current research also aims at identifying the timing of such mechanisms in order to assess which barriers appear early and initiate the speciation process, relative to those which appear later in the process but may 'finalise' speciation and prevent reverse speciation or breakdown on secondary

contact (Marie Curie Speciation Network, 2012). Incipient species may be the most useful for examining generation of reproductive isolation, where barriers that contributed to speciation (rather than arose after speciation was complete) can be tested (Coyne & Orr 2004; Via 2009; Seehausen *et al.* 2014; Puebla *et al.* 2014). Furthermore, testing the degree of differentiation or reproductive isolation between lineages at varying levels of divergence allows inference of the rate of isolation development, and assessment of which barriers occur early and remain throughout the speciation continuum (Coyne & Orr 2004; Merrill *et al.* 2011; Abbott *et al.* 2013). As standing variation has been shown to play an important role in local adaptation (e.g., Brawand *et al.* 2014), considering divergence at different timescales and spatial separation may be important in disentangling how mechanisms progress during divergence (Faria *et al.* 2014).

The advent of next generation sequencing (NGS) technology has allowed greater insight into the genetic basis for divergence-with-gene-flow mechanisms, by enabling the detection of heterogeneous levels of differentiation across the genome. Genomic scans and identification of outlier loci for population genetic measures of diversity (e.g., F_{ST}, D, D_{XY}, π) between lineages have become commonly used methods in the study of genomic patterns of speciation (e.g., Via & West 2008; Nosil *et al.* 2009; Hohenlohe *et al.* 2010; Jones *et al.* 2012; Brawand *et al.* 2014). Furthermore, the increasing availability of well-annotated reference genomes means that not only can regions of differentiation be localised on the genome, but also that the underlying genomic architecture of heterogeneous regions (that may resist recombination) can be investigated (e.g., Franchini *et al.* 2014; Fan & Meyer 2014). In the face of ongoing gene flow, in order for reproductive isolation to build up between incipient species, there must be 'barrier loci' which may be under divergent selection, or contribute to assortative mating (Abbott *et al.* 2013). As such, narrow regions of the genome that are highly divergent against a backdrop of low differentiation (outlier loci) are frequently interpreted as signals of adaptation or reproductive isolation, and have been referred to as 'islands of differentiation' (Turner *et al.* 2005; Turner & Hahn 2010; Renaut *et al.* 2011; Feder *et al.* 2012). However, the validity of identifying regions as outliers implicated in the speciation process based on statistical measures of allelic frequency or heterozygosity has been called in to question. In particular, the biases of relative measures of divergence (F_{ST}, Wright 1943; Da, Nei & Li 1979) compared to absolute measures (D_{XY}, Nei 1987) has been emphasised (Noor & Bennett 2009; Cruickshank & Hahn 2014). The use of relative measures of diversity may inflate patterns of differentiation because of their dependence on within-population variation

(Charlesworth 1998), and reanalysing 'islands of differentiation' identified with relative measures using absolute measures has indicated no more divergence than non-outlier regions of the genome (Cruickshank & Hahn 2014). However, relative measures may be better at identifying differentiation in very recent divergences (Noor & Bennett 2009; Cruickshank & Hahn 2014). Figure 4.1 illustrates how relative and absolute divergence measures compare in detecting differentiation dependent on divergence time (number of generations since lineage split) and levels of gene flow. As such, we may not expect D_{XY} to show substantial patterns of variation in recently diverged species exhibiting very low levels of differentiation.

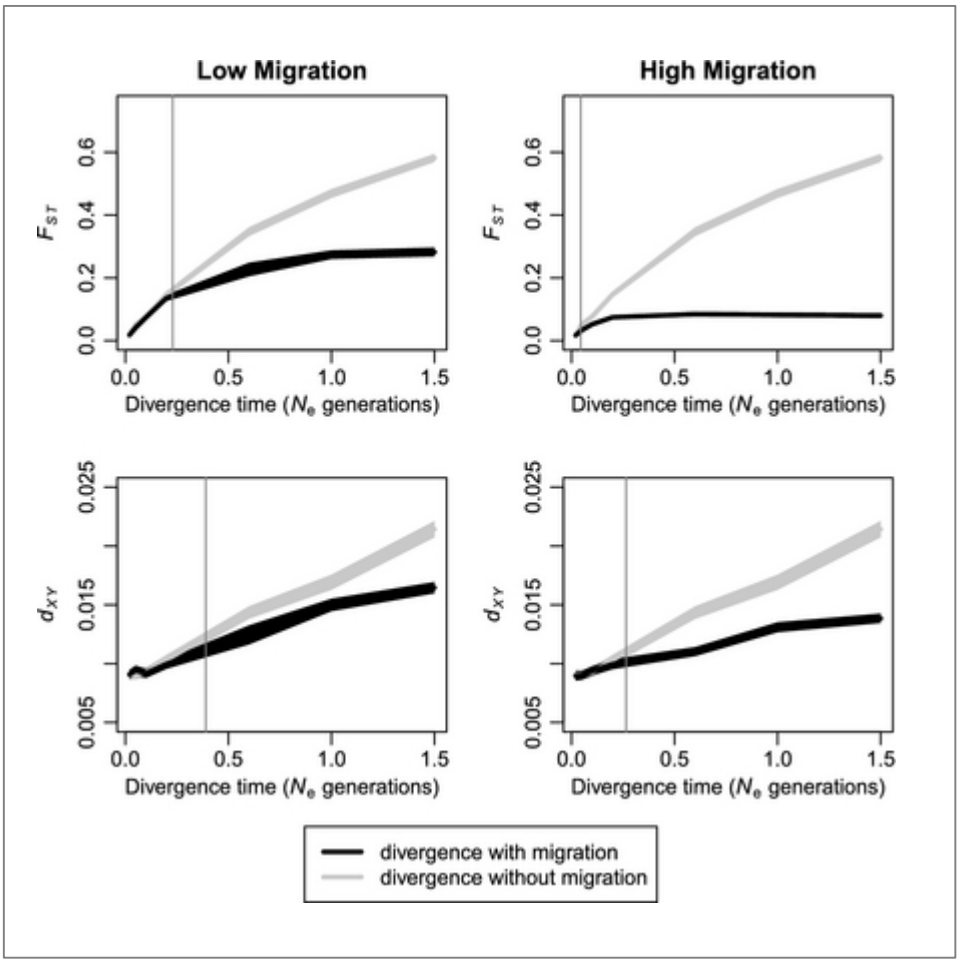


Figure 4.1 Comparison of relative and absolute measures of divergence.
Reproduced from Cruickshank *et al*. 2014 with permission from John Wiley and Sons. Simulated data sets of population divergence (grey and black lines, width=500 simulated dataset confidence interval) with either high or low migration. The point at which the two lines separate (grey vertical line) indicates the time (in generations) since the split at which each divergence measure is capable of detecting differentiation. In both cases, F_{ST} is able to detect divergence in the initial stages of separation (recent speciation) earlier than D_{XY}.

The influence of window size on detection of divergence has also been emphasised, with the use of small regions introducing biases in certain differentiation estimators (Martin *et al.* 2014), prompting the caveat that no one measure of divergence may be suitable for all comparisons. The detection of outliers may be as simple as visual inspection of genome-wide plots of F_{ST} (or other measures of divergence) to identify peaks of differentiation, or may employ an arbitrary cut-off point such as considering the loci exhibiting the highest (e.g., 1%, 5% or 10%) values as outliers. Other methods include the modelling of population genetic parameters (typically with an island-based model of gene flow) to identify loci that have comparably high or low values compared to neutral expectations: as computed in software programs including fdist (Beaumont & Nichols 1996), Lositan (Antao *et al.* 2008), and BayeScan (Foll & Gaggiotti 2008). Separate comparisons of several programs have indicated that BayeScan typically provides the lowest rate of error, particularly false positives (Perez-Figueroa *et al.* 2010; Narum & Hess 2011). Another way to address the issue of false positives (high estimates of F_{ST} resulting from processes other than divergent selection) is to investigate several populations, with the assumption that loci that are outliers across replicate populations are unlikely to have occurred by chance (Puebla *et al.* 2014). Although BayeScan was originally developed for detection of analysis between intraspecific populations, it has successfully been implemented in analysis of recent (incipient) species divergence, including in pupfish (Martin & Feinstein 2014) and hamlets (Puebla *et al.* 2014), as well as isolated lineages of salmon (Seeb *et al.* 2014). In addition to genomic scans to detect regions of differentiation, other research has considered phylogenetic discordance across the genome to consider the proportion that has been subject to introgression (Martin *et al.* 2013).

The *Alcolapia* radiation provides an excellent setting in which to conduct research on speciation processes, as it includes lineages and populations at differing levels of separation such that the processes involved at different stages of the speciation continuum can be examined. These comparisons include: allopatrically separated non-sister species (*A. grahami* and *A. latilabris*/*A. ndalalani*), sympatrically occurring non-sister species (*A. alcalica* and *A. latilabris*/*A. ndalalani*); sympatric sister species (*A. ndalalani* and *A. latilabris*); intraspecific allopatric populations of recent separation (*A. alcalica* north/south clades) and very recent separation (*A. grahami* Magadi/Nakuru lakes); intraspecific sympatric morphs of colour and trophic morphology (*A. alcalica* yellow/blue colour morphs and terminal/upturned mouth morph). Furthermore, the lack of other fish and cichlid species means that inference of ecological and/or sexual selection and competition

pressures is more straightforward in this system. This chapter employs several methods to investigate the distribution of differentiation across the genome and to test whether loci are under selection between species and populations of *Alcolapia*.

Aims

This chapter investigates genome-wide differentiation and whether there is a signal of selection between species, populations and morphs of the *Alcolapia* species radiation. The chapter specifically seeks to address the following questions: i) are there regions of the genome under diversifying selection between lineages (both in species occurring sympatrically, and within species between different environments)? ii) do genomic regions of high differentiation correspond to differences in phenotype between species and morphs? iii) is there evidence that few genomic regions of large effect contribute to early stage divergence? iv) given the frequent interplay of ecological and sexual selection (e.g., sensory drive), is there evidence for differentiation between sexes in an ecological context?

Methods

RAD dataset

The current chapter provides additional analysis of the RAD dataset produced following the methods discussed in chapter three. For these analyses, only the reference-aligned sequences were used (datasets A and C from chapter three). Where necessary, subsets of these datasets were produced using custom bash scripts to extract specific genomic regions.

Phylogenomic inference

All phylogenomic analysis was conducted using the SSE PTHREADS version of RAxML 8, employing the GTRGAMMA model.

Sliding window F_{ST} and D_{XY}

Sliding window F_{ST} analyses were conducted between species, populations, morphs, and sexes in the EggLib Python module (De Mita & Siol 2012) using custom python scripts. Analyses were run on the entire set of *Alcolapia* filtered

biallelic SNPs (91 individuals; 22.2 Mb) with pre-defined populations, a window size of 1 Mb, a slide length of 100,000 bp, and included only windows with a minimum of 10,000 sites, excluding unplaced scaffolds. Populations were pre-defined and individuals included selected to ensure even numbers in each comparison. For species analyses, numbers were constrained to eight individuals (*Alcolapia*) as the maximum sample number of *A. grahami* from Lake Magadi, or three samples (*O. amphimelas*), as the maximum sample number from Lake Manyara. Where more than the pre-defined number of individuals were available, selection was made based on geographic sampling and sequence quality of the RAD data. Species comparisons that included *A. alcalica* were further divided in geographical subsections, including 'southern' (sites 5 and 11) and 'northern' (sites 6 and 19) populations as identified by the phylogenomic analysis in chapter three. Sliding window analysis was conducted for the population comparisons listed in Table 4.1. Only individuals that could be sexed by dissection were included for sex comparisons – any specimens with undifferentiated gonads (eight individuals) were excluded.

Table 4.1. Comparisons included in BayeScan and sliding-window analyses.

		Comparator A	Comparator B
Species	A	*A. alcalica* "north" (n=8)	*A. latilabris* (n=8)
	B	*A. alcalica* "south" (n=8)	*A. latilabris* (n=8)
	C	*A. alcalica* "north" (n=8)	*A. ndalalani* (n=8)
	D	*A. alcalica* "south" (n=8)	*A. ndalalani* (n=8)
	E	*A. alcalica* "north" (n=8)	*A. grahami* (n=8)
	F	*A. alcalica* "south" (n=8)	*A. grahami* (n=8)
	G	*A. alcalica* "north" (n=7)	*O. amphimelas* (n=7)
	H	*A. alcalica* "south" (n=7)	*O. amphimelas* (n=7)
	I	*A. latilabris* (n=8)	*A. ndalalani* (n=8)
	J	*A. latilabris* (n=8)	*A. grahami* (n=8)
	K	*A. ndalalani* (n=8)	*A. grahami* (n=8)
Population	L	*A. alcalica* "north" (n=8)	*A. alcalica* "south" (n=8)
	M	*A. alcalica* "north" males (n=6)	*A. alcalica* "south" males (n=6)
	N	*A. alcalica* terminal mouth (n=4)	*A. alcalica* upturned mouth (n=4)
	O	*A. alcalica* blue morph (n=4)	*A. alcalica* yellow morph (n=4)
	P	*A. grahami* Magadi (n=4)	*A. grahami* Nakuru (n=4)
	Q	*O. amphimelas* Eyasi (n=3)	*O. amphimelas* Manyara (n=3)
Sex	R	*Alcolapia* male (n=15)	*Alcolapia* female (n=15)
	S	*A. alcalica* male (n=3)	*A. alcalica* female (n=3)
	T	*A. latilabris* male (n=3)	*A. latilabris* female (n=3)
	U	*A. ndalalani* male (n=4)	*A. ndalalani* female (n=4)
	V	*A. grahami* male (n=4)	*A. grahami* female (n=4)

For a subset of species-level comparisons (B, D, F, H, and I from Table 4.1), analyses were conducted to assess whether the top 1% and 5% F_{ST} outliers were distributed non-randomly across the genome. For these tests, F_{ST} was calculated in non-overlapping windows (to preclude non-independence of windows), using a window size of 100Kb, a slide length of 100 Kb, and a minimum of 1,000 sites. The non-random distribution of outlier windows was tested using 10,000 permutations, comparing the closest inter-peak distances between observed and permuted datasets, and employing a nearest neighbour index (NNI) as an indicator of the level of clustering in the observed data (Clark & Evans 1954). A modified NNI ratio was calculated, which used the mean of the permuted data as the ratio denominator (rather than the standard random-distribution denominator of points/distance), to avoid the assumption of a purely linear genome, and for which only distances within linkage groups were calculated. A Z-statistic was used to test whether the modified NNI was significantly different from the mean random distribution (Clark & Evans 1954; Hammond & McCullagh 1978). Bin numbers for the distance calculations (for each non-overlapping window) were used rather than taking a midpoint chromosomal bp location, although this had no effect on the significance of the

results (tested in 50% of comparisons). All permutations and significance testing were conducted in R 3.1.2 (R Core Team 2014).

Outlier loci

Detection of loci under selection was conducted by looking for F_{ST} outlier loci implementing a Bayesian approach in BayeScan 2.1 (Foll & Gaggiotti 2008) between the same populations as for the sliding-window analyses. BayeScan decomposes F_{ST} into locus- (alpha) and population-specific (beta) components and implements a reversible-jump MCMC algorithm to estimate the posterior distribution of models with and without selection. Selection is inferred when alpha is significantly different from zero and the beta component is insufficient to explain differentiation. BayeScan has been shown to produce a high number of false positives in cases of isolation by distance or range expansion (Lotterhos & Whitlock 2014), although neither is thought to be a factor in the present analysis, given the lack of evidence of isolation by distance (tested in chapter three) and the restricted range size of all populations. BayeScan has been reported to be conservative with respect to other methods for detecting outlier loci (Henry & Russello 2013). For these analyses, a minimum allele frequency of 10% was imposed, with a missing data threshold at each site of 25% across all individuals in each comparison. Input files were formatted using PGDSpider 2.0.8.0 (Lischer & Excoffier 2012), and prior odds for the neutral model were set at 10, using default parameters for the MCMC analysis. A single comparison (*A. alcalica* terminal and upturned mouth morphs) was repeated at prior odds of 1 and 0.1 to consider the effects of the prior odds on outlier detection. All results were analysed to identify outlier loci at false discovery rates of FDR=0.10 and FDR=0.05.

Linkage disequilibrium

Where potential sites of selection (i.e., increased differentiation between populations relative to the rest of the genome) were identified, linkage disequilibrium around these sites was investigated using Haploview 4.2 (Barrett *et al.* 2005) to calculate linkage disequilibrium measures Dprime (D') and r^2. These analyses were conducted on individual linkage groups based on results from the sliding window and BayeScan analyses. Input files were created containing only the individuals for each comparison and the linkage group of interest, with a maximum missing data threshold of 25% and a minimum allele frequency of 0.2.

Heterozygosity

Given the fragmentary nature of the soda lake environment, heterozygosity was tested to see if isolation led to a loss of heterozygosity in any of the *Alcolapia* populations. Estimates of heterozygosity using SNPs have been shown to be as robust as estimates using microsatellites, and accuracy improved with increasing marker number (Miller *et al.* 2013). Individual heterozygosity estimates were calculated as the number of heterozygous sites divided by total number of bases called per individual (Hoffman *et al.* 2014). Calculations were run on the full alignment and a dataset with no missing data (i.e., including only sites for which there were data for all individuals).

Results

Outlier loci

A subset of species-level comparisons were first compared using sliding window analyses of F_{ST} and D_{XY} to consider the differences in relative and absolute measures of divergence between species. The sliding-window analyses of F_{ST} indicated heterogeneous differentiation across the genome in the *Alcolapia* comparisons, with several peaks of divergence in each pairwise comparison against a background of low divergence (Figure 4.2). This is in contrast to the *A. alcalica*/*O. amphimelas* comparison, which exhibited uniformly high values of F_{ST} across the genome. The D_{XY} analyses showed less substantial variation across the genome, and fewer peaks of high diversity, but the highest peak was found in all within-*Alcolapia* comparisons on linkage group 23 (Figure 4.2), although this peak was not identified by the BayeScan analysis (see below). The LG23 peak in D_{XY} corresponded to a reduction of F_{ST} at the same region in the *A. alcalica/O. amphimelas* comparison (Figure 4.2)

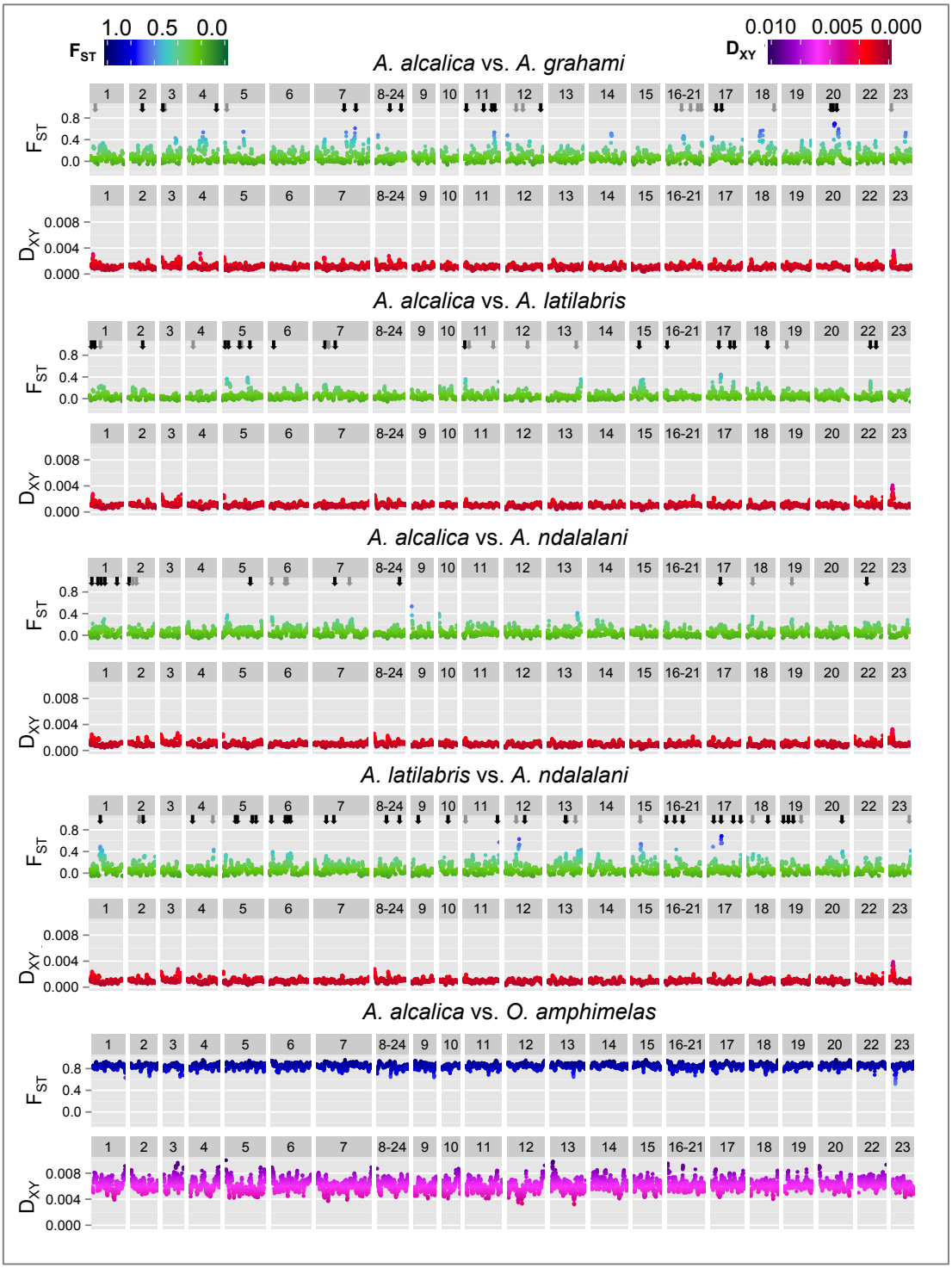


Figure 4.2. Sliding-window analysis of relative (F_{ST}) and absolute (D_{XY}) divergence for pairwise species comparisons.

Results are plotted by linkage group of the reference genome (*O. niloticus*) as indicated by numbers in the upper grey bar of each plot. Window size is 1Mb with a slide of 100kb. Approximate genome position of F_{ST} outliers identified by BayeScan (Table 4.2) are indicated by arrows for FDR=0.05 (black) and FDR=0.10 (grey).

Plotting the frequency distribution of the sliding-windows (Figure 4.3), exhibited a right-skewed pattern for within-*Alcolapia* comparisons with a majority of windows showing low differentiation, but a small number showing comparatively high F_{ST} values. Conversely, the *Alcolapia*-outgroup comparison showed a left-skewed distribution, with most comparisons showing high levels of differentiation and only a few regions of low differentiation.

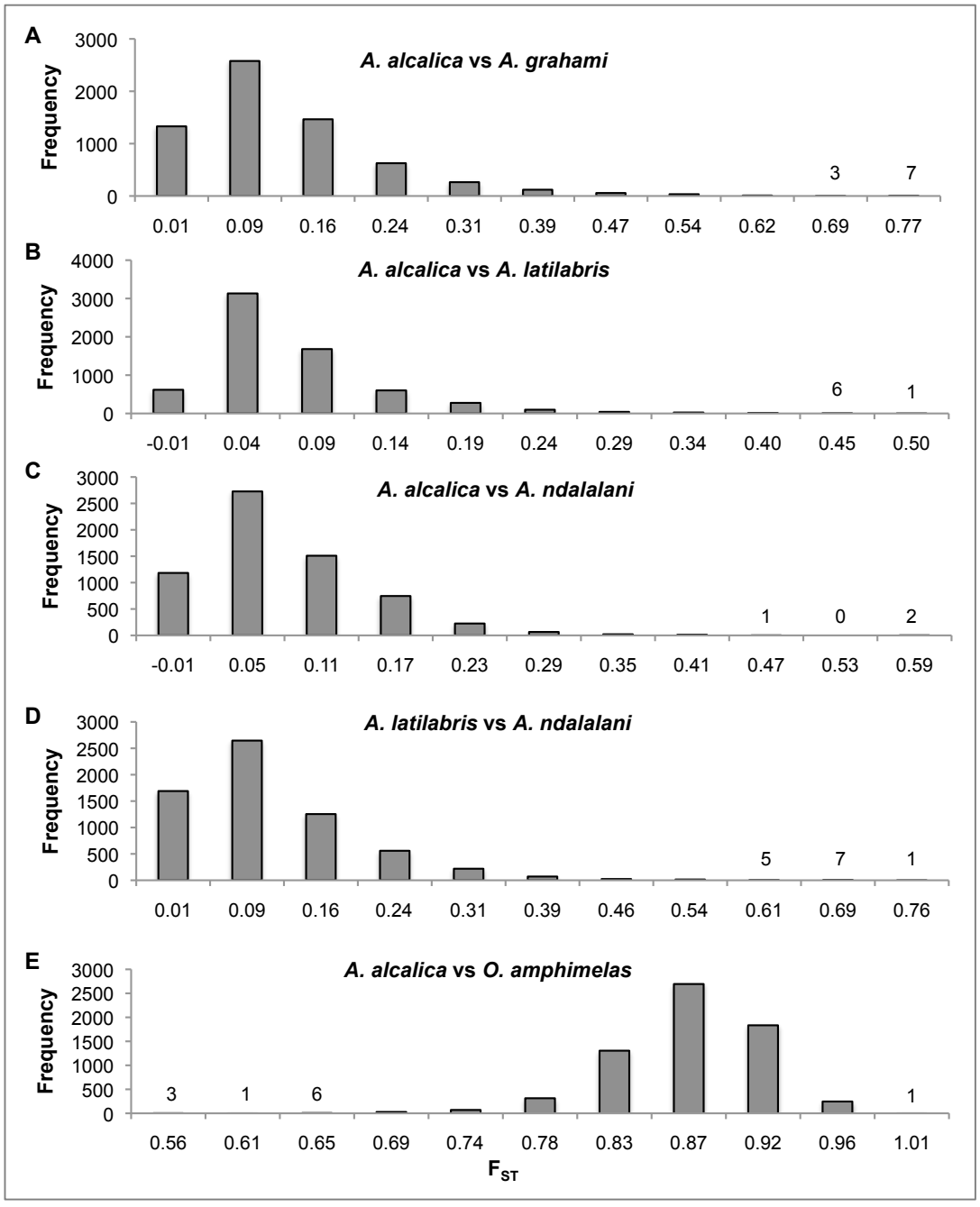


Figure 4.3. Frequency histograms of sliding-window F_{ST}.

F_{ST} values from the analysis in Figure 4.2 binned in 10 equal-width bins across the range of values, with the x-axis categories being the upper limit of each bin. Frequency values <10 are given in text above respective columns.

The F_{ST} values for non-overlapping windows were also calculated, and plotting the top 1% and 5% F_{ST} windows in the genome-wide analysis indicated heterogeneous distribution across the genome (Figure 4.4). The range of F_{ST} values covered by the top 5% of values was considerably larger within the *Alcolapia* comparisons (F_{ST}=0.2-0.8) than in the *Alcolapia*-outgroup comparison (F_{ST} =0.98-1.00).

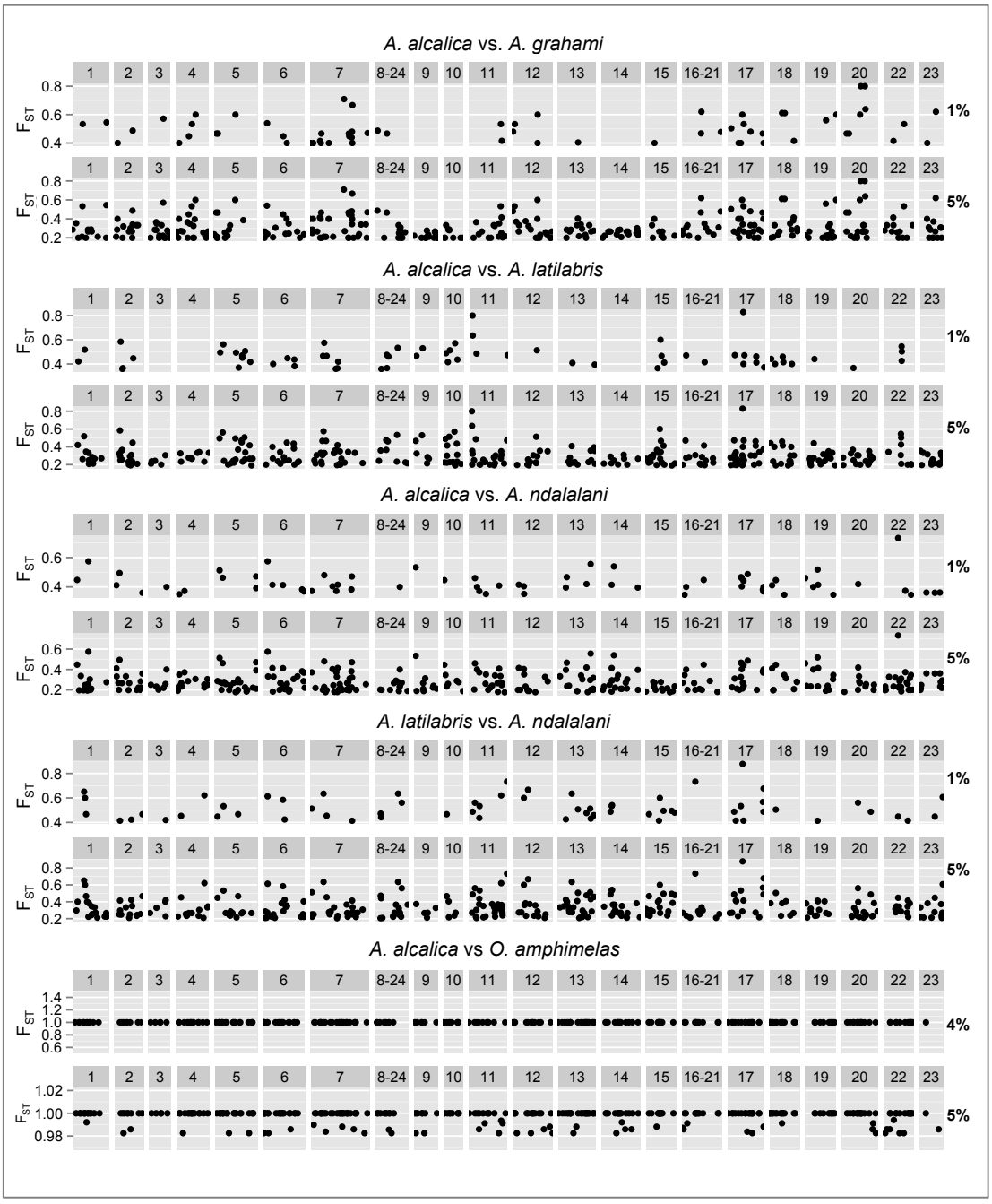


Figure 4.4. Plots of the top 1% and 5% of F_{ST} non-overlapping windows. The highest-scoring windows in the outgroup (*O. amphimelas*) comparison all exhibited maximal values (F_{ST}=1.00), so all windows are shown at this value (equivalent to 4% of windows) rather than 1%.

The distribution of 1% outliers was significantly non-random only in the *A. alcalica* vs. *A. latilabris* comparison (significant by permutation testing and Z-statistic of the NNI ratio), however all comparisons exhibited NNI<1, indicating tendency to clustering rather than dispersion.

Given the low levels of differentiation exhibited by D_{XY} in the initial species comparisons, only F_{ST} was investigated in the sliding-window analyses of additional comparisons for populations and morphs. Comparison of the northern and southern *A. alcalica* clades to other species suggested that regions under selection are the same for both clades (Figure 4.5). Generally, the same peaks are present between both comparisons, however for the comparisons to *A. latilabris* and *A. ndalalani*, the peaks are mostly reduced in pairwise comparisons with the *A. alcalica* southern clade, suggesting that the homogenising effects of gene flow restrict the differentiation of these peaks relative to the northern clade. Identifying peaks that are of comparable differentiation extent between the northern and southern clades (e.g., peaks on linkage groups 9 and 10 in the *A. ndalalani* comparison to northern and southern clades, Figure 4.5) may indicate regions of differentiation that are not reduced by gene flow. The levels of differentiation between *A. alcalica* and *A. grahami* are comparable between both northern and southern *A. alcalica* clades. Most peaks are shared, and there is not a clear trend of either comparison exhibiting higher levels of differentiation. Similarly, the high levels of differentiation from *O. amphimelas* is comparable between southern and northern *A. alcalica* clades, as would be expected given the increased time since the divergence from the outgroup relative to divergence within *Alcolapia*.

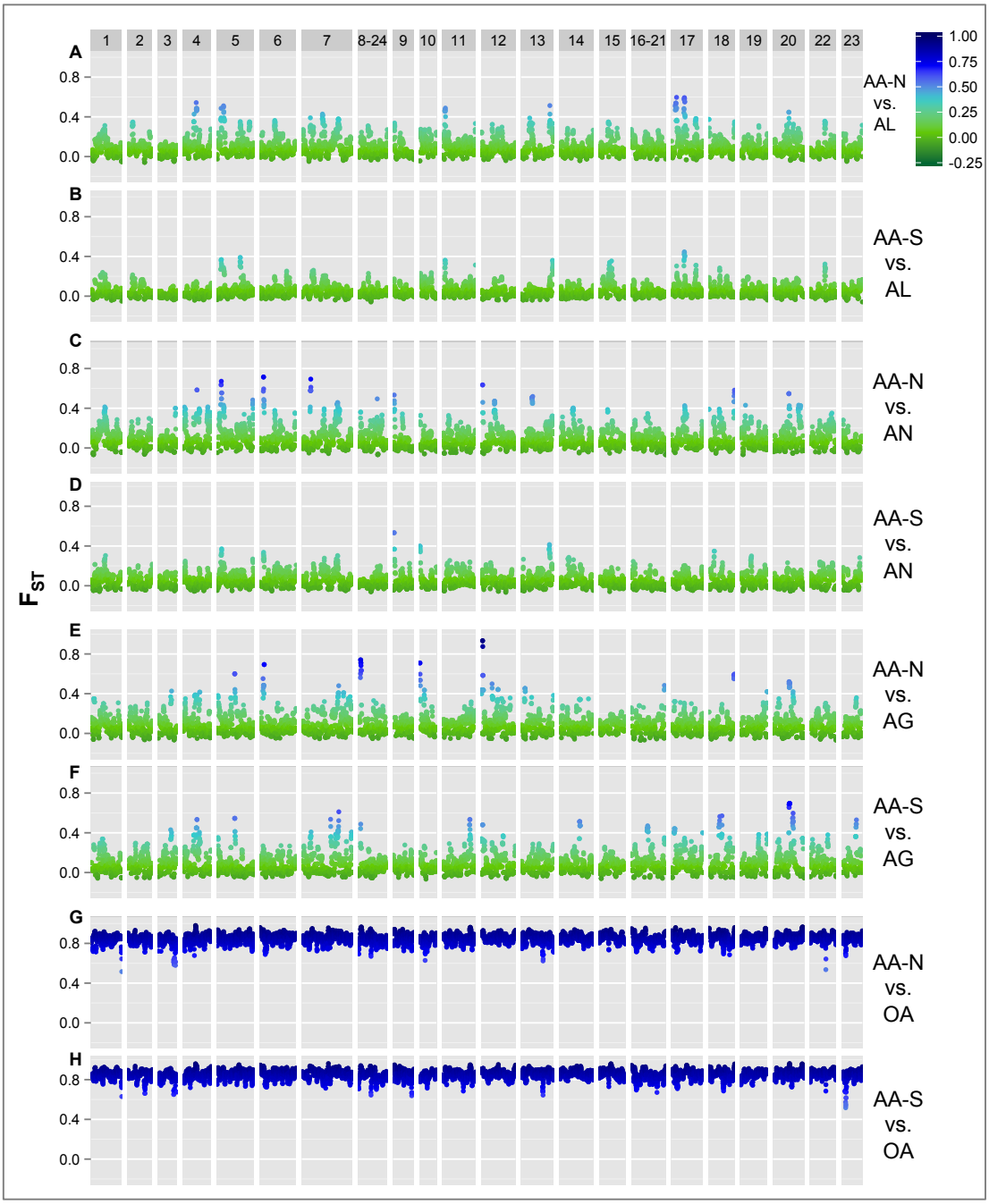


Figure 4.5. Sliding window F_{ST} analysis for *A. alcalica* clades vs. other species.
Letters at the top-left hand corner in each plot indicate comparisons as listed in Table 4.1. The comparison for each plot is also given on the right-hand side of each plot. AA-N: *A. alcalica* northern clade; AA-S: *A. alcalica* southern clade; AL: *A. latilabris*; AN: *A. ndalalani*; AG: *A. grahami*; OA: *O. amphimelas*.

As may be expected from the phylogenomic results in chapter three, comparison of the three Natron species revealed low levels of differentiation with only a few peaks between *A. latilabris* and *A. ndalalani*, but considerably higher levels of differentiation between these two species and *A. grahami* (Figure 4.6). Furthermore,

the peaks are generally shared in the latter two comparisons (though appear slightly more pronounced in the *A. latilabris* comparison).

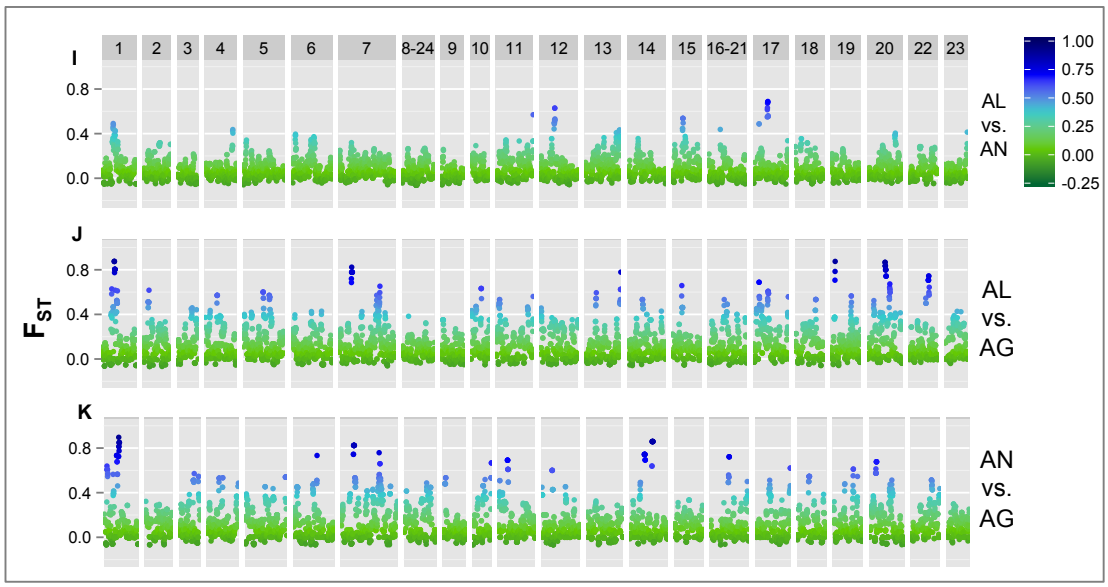


Figure 4.6. Sliding window F_{ST} analysis of additional species comparisons.
Letters in left-hand corner indicate comparisons listed in Table 4.1. AL: *A. latilabris*; AN: *A. ndalalani*; AG: *A. grahami*.

Intraspecific comparisons unsurprisingly showed lower levels of differentiation than the species level comparisons (Figure 4.7). Comparing the northern and southern *A. alcalica* clades exhibited low levels of differentiation, and although several small peaks were noticeable all were F_{ST}<0.4. Excluding the female specimens from the comparison and comparing only males did not appear to have any effect on the results (Figure 4.7; plots L and M). Meanwhile the comparison of terminal and upturned *A. alcalica* morphs at site 15 exhibited two clear peaks on LG7 (F_{ST}=0.5) and LG11 (F_{ST}~0.8), with several smaller peaks around F_{ST}~0.4. Furthermore the larger peak on LG11 exhibited wide shoulders, with several windows around the main peak exhibiting elevated F_{ST}, such that no windows in this region seemed to exhibit the baseline level of low differentiation (Figure 4.7; plot N). The intraspecific yellow and blue *A. alcalica* morphs exhibited lower levels of differentiation, but with three peaks (F_{ST}=0.2-0.4) on LG1, LG5 and LG17 (Figure 4.7; plot O). Surprisingly, given the very short time since allopatric separation (~60 years), the comparison of Lake Magadi and Lake Nakuru populations also exhibited peaks of differentiation (Figure 4.7; plot P). Finally the *O. amphimelas* comparison revealed highly variable levels of differentiation (F_{ST}=0-1) between Lake Manyara and Eyasi populations, however most of the genome exhibited high levels of differentiation, with only a few

regions of much reduced differentiation (LG8-24 and part of LG13; Figure 4.7, plot Q).

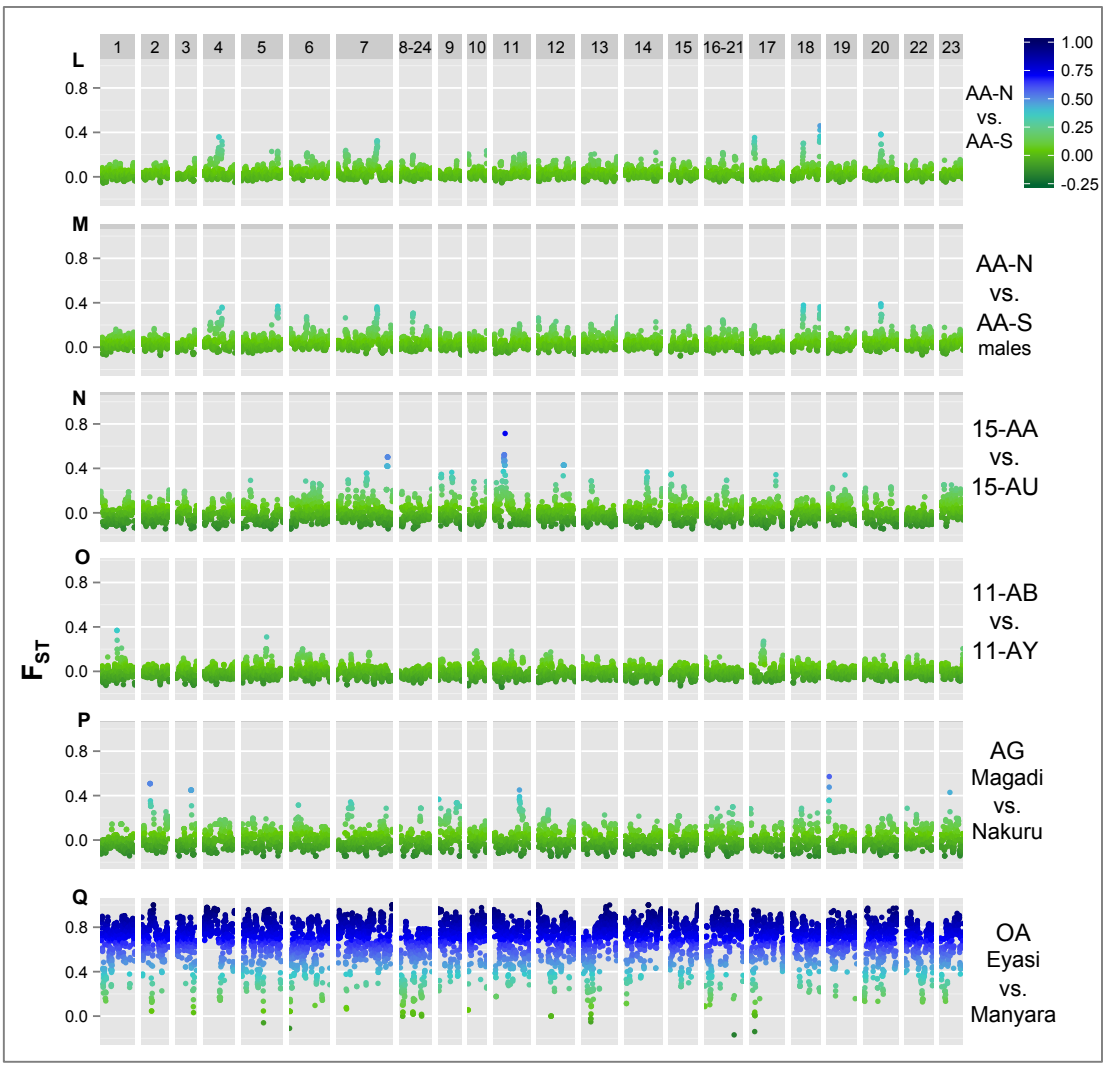


Figure 4.7. Sliding window F_{ST} analysis of intraspecific comparisons.
Letters in left hand corner of each plot refer to the comparisons in Table 4.1, also specified on the right-hand side of each plot. AA-N: *A. alcalica* northern clade; AA-S: *A. alcalica* southern clade; 15-AA: site 15 *A. alcalica* terminal-mouth morph; 15-AU: site 15 *A. alcalica* upturned-mouth morph; 11-AB: site 11 *A. alcalica* blue morph; 11-AY: *A. alcalica* yellow morph. AG: *A. grahami*; OA: *O. amphimelas*.

Comparisons between sexes did not reveal any peaks of differentiation across the entire *Alcolapia* radiation, suggesting no common differentiation of sexes across all species (Figure 4.8; plot R). Comparison of sex within species also did not indicate clearly differentiated peaks, other than possibly for the *A. grahami* comparison on LG7 (Figure 4.8). However, it should be noted that due to (unintentional) male-biased sampling, numbers were low for these species comparisons, so may have

insufficient power to detect signals of selection (Table 4.1 for sample numbers per comparison). Furthermore, association tests for sex-linked markers (SNPs with high homozygosity in either sex, or SNPs with x2 coverage in either sex) conducted by the sequencing facility (Edinburgh Genomics) did not yield any candidates (data not shown).

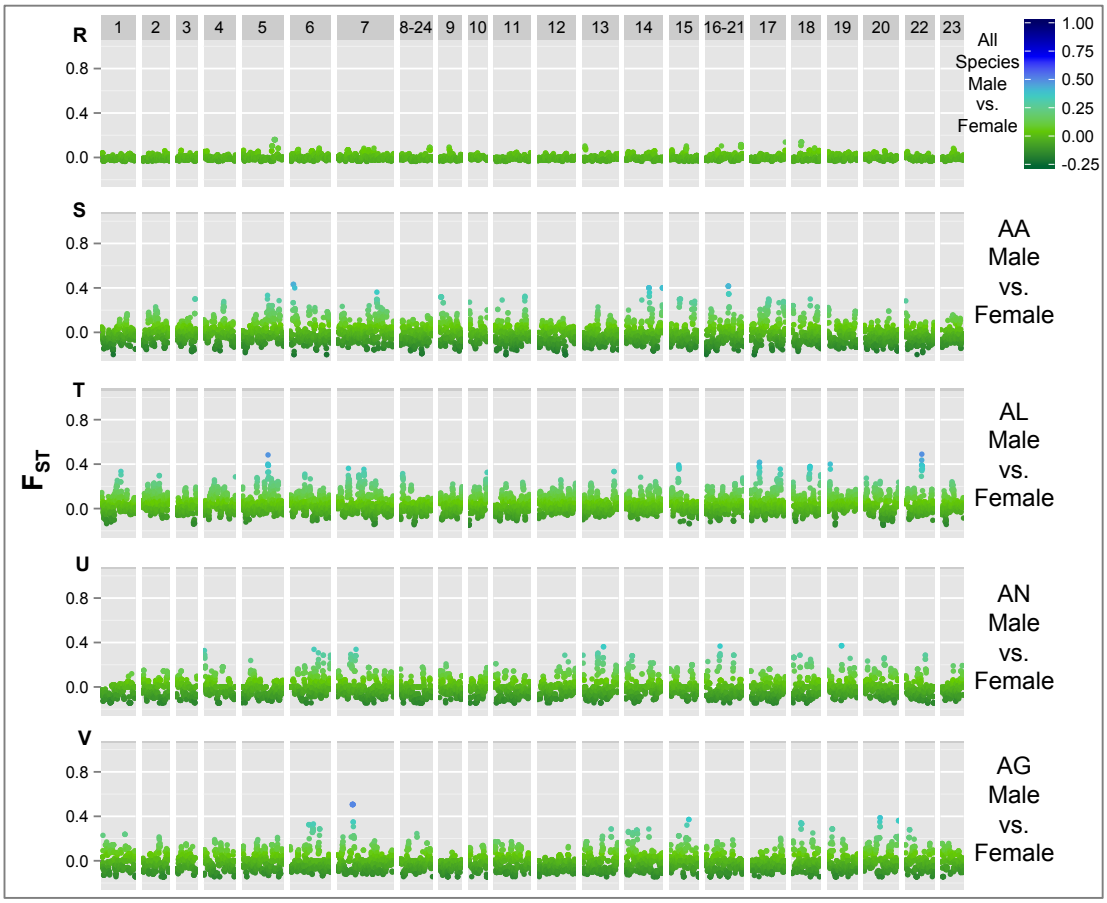


Figure 4.8. Sliding window F_{ST} analyses of sex within species.
AA*: A. alcalica*; AL: *A. latilabris*; AN: *A. ndalalani*; AG: *A. grahami*.

For species level comparisons, results of BayeScan analysis to detect outlier loci were generally congruent with the sliding window analyses. Comparisons with the northern *A. alcalica* clade exhibited a higher number of outliers compared to those with the southern clade (Table 4.2). The comparison of Natron species to *A. grahami* also exhibited similar patterns, with the *A. latilabris* comparison detecting more outliers than the *A. ndalalani* comparison. Perhaps surprisingly, given the more recent divergence time and closer phylogenetic relationship, the *A. alcalica* northern clade exhibited higher number of outliers than *A. grahami* in comparisons with *A. latilabris* and *A. ndalalani* (Table 4.2). All of the outliers indicated diversifying selection (alpha > 0) rather than balancing selection, and the majority of outliers had

$\log_{10}$ Bayes Factor scores > 0.5 (considered substantial on Jeffrey's scale of evidence; Foll 2012). For each comparison with the *A. alcalica* southern clade, these were: *A. alcalica*/*A. grahami*: 86%; *A. alcalica*/*A. latilabris*: 87%; *A. alcalica*/*A. ndalalani*: 100%; *A. latilabris*/*A. ndalalani*: 96%. Neither of the species comparisons with *O. amphimelas* detected outliers in comparison with *A. alcalica* northern or southern clades (even at the relaxed false discovery rate FDR=0.1). While these species comparisons are in line with the sliding window analyses, other than the comparison of *A. alcalica* northern and southern clades, none of the intraspecific morph/population/sex comparisons yielded any outlier loci. These comparisons had the fewest individuals included (except the sex comparison across all species), which may make detection of differentiation difficult (although low sample numbers are normally expected to inflate estimates of F_{ST}; Nadeau *et al.* 2012a). Replicate runs of comparisons yielded similar results across all runs.

BayeScan default settings are considered fairly conservative (Foll 2012). The default prior odds setting of 10 indicates the relative likelihood of the neutral model over the selection model (i.e., 10 times as likely), and increasing the odds value should give a lower level of false estimates, which may be appropriate for large datasets (Lotterhos & Whitlock 2014). However, the default value of 10 has been reported to give proportion of outlier loci of up to 30% in some datasets (Cullingham *et al.* 2014), whereas analysis on the *Alcolapia* species identified <0.5% of outlier loci in all comparisons (Table 4.2), which is substantially less than the 2-10% of loci thought to underlie ecologically relevant traits (Stinchcombe & Hoekstra 2007). As such, one of the intraspecific morph analyses (terminal and upturned mouth morphs was repeated with lower prior odds values of 1 and 0.1 to see if this increased the detection of outlier loci. However, runs at the decreased prior odds values also did not yield any outlier loci.

Table 4.2. F_{ST} outliers.

Outliers were identified using Bayescan (Foll & Gaggiotti 2008) using false discovery rates (FDR) of 0.05 and 0.10.

All outliers identified in each comparison exhibited alpha > 0, indicative of diversifying selection.

				All SNPs		Outliers FDR=0.05		Outliers FDR=0.10	
		Comparator A	**Comparator B**	**N**	**F_{ST} range**	**n (%)**	**F_{ST} range**	**n (%)**	**F_{ST} range**
Species	A	*A. alcalica* “north”	*A. latilabris*	27,309	0.08-0.47	119 (0.44)	0.27-0.47	158 (0.58)	0.22-0.47
	B	*A. alcalica* “south”	*A. latilabris*	30,841	0.04-0.35	39 (0.13)	0.16-0.35	55 (0.18)	0.15-0.35
	C	*A. alcalica* “north”	*A. ndalalani*	23,939	0.08-0.47	100 (0.42)	0.24-0.47	139 (0.58)	0.28-0.47
	D	*A. alcalica* “south”	*A. ndalalani*	28,026	0.04-0.28	14 (0.05)	0.18-0.28	27 (0.10)	0.16-0.28
	E	*A. alcalica* “north”	*A. grahami*	17,617	0.09-0.49	40 (0.23)	0.29-0.49	52 (0.30)	0.26-0.49
	F	*A. alcalica* “south”	*A. grahami*	23,264	0.01-0.49	31 (0.13)	0.32-0.49	45 (0.19)	0.27-0.49
	G	*A. alcalica* “north”	*O. amphimelas*	71,076	0.68-0.85	0	-	0	-
	H	*A. alcalica* “south”	*O. amphimelas*	82,474	0.68-0.85	0	-	0	-
	I	*A. latilabris*	*A. ndalalani*	22,946	0.05-0.42	43 (0.19)	0.22-0.42	75 (0.33)	0.18-0.42
	J	*A. latilabris*	*A. grahami*	18,434	0.13-0.53	76 (0.41)	0.38-0.53	108 (0.59)	0.33-0.53
	K	*A. ndalalani*	*A. grahami*	15,018	0.14-0.54	58 (0.38)	0.39-0.54	81 (0.54)	0.35-0.54
Population	L	*A. alcalica* “north”	*A. alcalica* “south”	33,283	0.03-0.30	11 (0.03)	0.18-0.30	25 (0.08)	0.16-0.30
	M	*A. alcalica* “north” males	*A. alcalica* “south” males	37,448	0.03-0.30	6 (0.02)	0.21-0.30	18 (0.05)	0.17-0.30
	N	*A. alcalica* terminal mouth	*A. alcalica* upturned mouth	17,400	0.0007-0.0008	0	-	0	-
	O	*A. alcalica* blue morph	*A. alcalica* yellow morph	34,512	0.004-0.004	0	-	0	-
	P	*A. grahami* Magadi	*A. grahami* Nakuru	11,448	0.0012-0.0013	0	-	0	-
	Q	*O. amphimelas* Eyasi	*O. amphimelas* Manyara	23,349	0.52-0.56	0	-	0	-
Sex	R	*Alcolapia* male	*Alcolapia* female	13,373	0.0003-0.0004	0	-	0	-
	S	*A. alcalica* male	*A. alcalica* female	16,390	0.0008-0.0009	0	-	0	-
	T	*A. latilabris* male	*A. latilabris* female	27,714	0.006-0.008	0	-	0	-
	U	*A. ndalalani* male	*A. ndalalani* female	10,521	0.0009-0.001	0	-	0	-
	V	*A. grahami* male	*A. grahami* female	18,042	0.0007-0.0007	0	-	0	-

Phylogenomic incongruence

Given the results of the sliding-window analysis of D_{XY} (Figure 4.2), which exhibited a peak on LG23 across all comparisons, phylogenomic analysis was conducted on the region of the peak to see if the divergence corresponded to phylogenetic signal. RAxML analysis was conducted on the entire linkage group (LG23) and another linkage group not displaying any elevated D_{XY} (LG22). Analysis was also conducted on the specific region of the D_{XY} peak (3 Mb along the linkage group, which contained 74,173bp of sequence data), and an adjacent non-peak region of the same chromosomal distance (3 Mb; 122,914 bp). Because the non-peak region contained substantially more sequence coverage, a second non-peak region was also tested (2 Mb; 79,657 bp). The analysis of whole linkage groups was congruent, with both LG22 and LG23 exhibiting similar topologies to the analysis of all linkage groups (chapter three), but with much reduced support values (Appendix Figure 4A.1). The subset analyses exhibited very little phylogenetic signal, with minimal bootstrap support and individuals not clustering by species or population. However, the tree generated from the region of the D_{XY} peak had substantially longer branch lengths than the non-peak regions, suggesting that the peak in D_{XY} represents an elevated diversity within this region, but it does not partition within species or populations, and does not contain phylogenetic signal (Appendix Figure 4A.1 and 4A.2). Furthermore, in the peak region, the branch length from *Alcolapia* to *O. amphimelas* is considerably shorter, while that of *O. niloticus* is much longer.

Linkage disequilibrium

Patterns of linkage disequilibrium around F_{ST} peaks were considered for two intraspecific comparisons of *A. alcalica* terminal vs. upturned mouth morphs and yellow vs. blue colour morphs. Linkage plots were calculated in Haploview for linkage groups exhibiting F_{ST} peaks. Linkage plots did not indicate increased linkage disequilibrium in regions of F_{ST} peaks (Appendix Figure 4A.3), as may have been expected if regions of these parts of the genome were resistant to recombination. Given the signal of differentiation from F_{ST} sliding window analysis, it is perhaps surprising that no signal of outlier loci (Table 4.2) or elevated linkage disequilibrium is found in these regions. It is possible that the use of different filtering parameters affects the analysis, as sliding window analyses were conducted on the unfiltered dataset, while BayeScan and LD were calculated on biallelic datasets filtered for

missing data and minimum allele frequency, and this affected the total number of SNPs in potential outlier loci between the comparisons (Table 4.3).

Table 4.3. Number of SNPs per region of interest in filtered datasets for different analyses.

Comparison	Linkage group	FST window position	window size (Mb)	SNPs (F_{ST})	SNPs (LD)	SNPs (BayeScan)
AY-AB	LG1	14300000-15400000	1.1	35	9	10
AY-AB	LG5	22400000-23400000	1	82	17	21
AY-AB	LG17	12000000-13300000	1.3	187	42	72
AA-AU	LG7	44900000-46700000	1.8	92	19	20
AA-AU	LG11	10400000-11400000	1	39	5	7
AA-AU	LG11	9100000-11400000	2.3	102	14	22
AA-AU	LG12	23500000-25200000	1.7	68	11	8

AY: *A. alcalica* yellow morph; AB: *A. alcalica* blue morph; AA: *A. alcalica* terminal mouth morph; AU: *A. alcalica* upturned mouth morph.

Heterozygosity

Genome-wide heterozygosity was calculated as the number of heterozygous sites per individual divided by number of total sites genotyped. Calculating the estimate across all genotyped sites (mean=13 Mb per individual) produced slightly higher estimates than calculating only for sites for which all individuals were genotyped (i.e., no missing data across all 84 individuals; 500 Kb) (Figure 4.9). For the dataset with no missing data, a limited number of species comparisons were significantly different in heterozygosity (Tukey's HSD; $P<0.05$): *A. grahami* site 21 vs. all *A. ndalalani* populations, all *A. latilabris* populations and *A. alcalica* sites 5, 15, 17; northern *A. alcalica* (6, 9) vs. southern *A. ndalalani* (5). *Alcolapia latilabris* and *A. ndalalani* generally showed lower heterozygosity than *A. alcalica* and *A. grahami* (Figure 4.9). None of the intraspecific comparisons were significant, suggesting that the isolation of certain populations (in particular the northern clade of *A. alcalica*) has not contributed to a loss of heterozygosity.

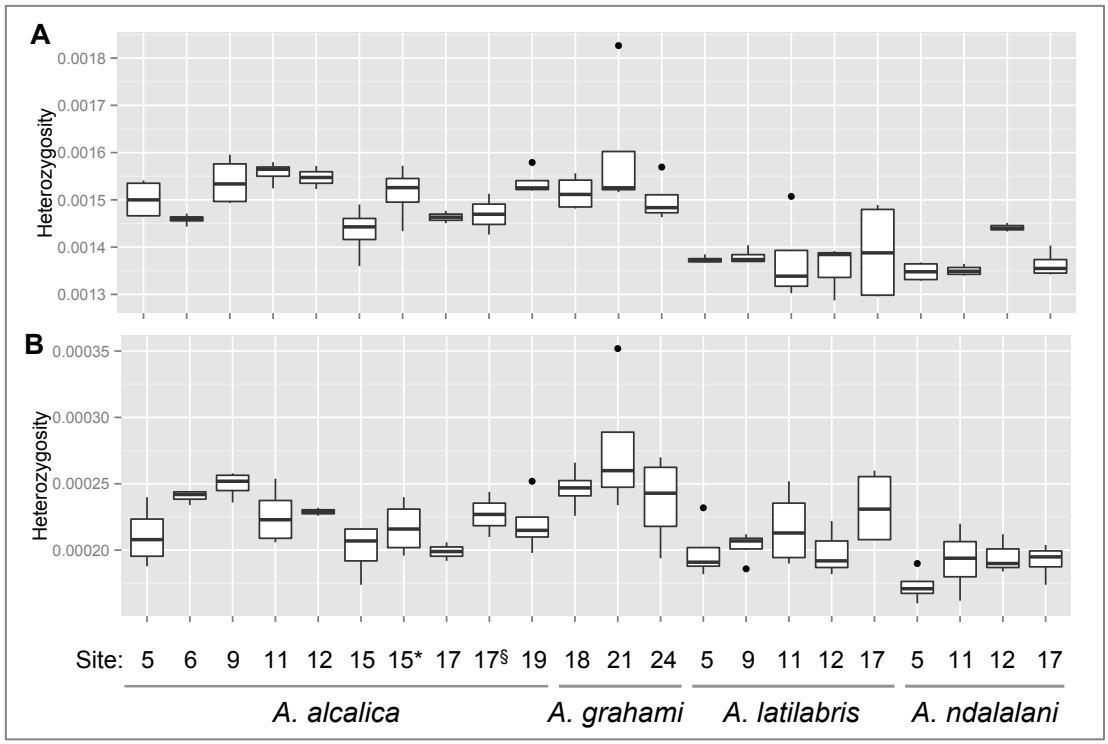


Figure 4.9. Heterozygosity of *Alcolapia* populations.

A) Calculated across all genotyped sites (mean: 13 Mb per individual). B) Calculated only across sites for which data were available for all 84 individuals (500 Kb). *indicates *A. alcalica* site 15 trophic morph (upturned mouth). §indicates site 17 *A.* aff. *ndalalani* identified as possible hybrids in phylogenomic analysis of chapter three.

Discussion

Genomic islands of differentiation in the *Alcolapia* species flock

Despite the low overall genomic differentiation between *Alcolapia* species (chapter three), the sliding window analyses support peaks of differentiation across the genome between species (Figures 4.2, 4.5, 4.6). Sliding-window analyses identified several high-F_{ST} windows distributed heterogeneously across the genome in all *Alcolapia* comparisons (Figures 4.2-4.4), and which were higher in comparisons of the *A. alcalica* northern clade than the southern clade. This pattern was also observed in the identification of several F_{ST} outliers in the BayeScan analysis, all of which indicated diversifying selection rather than balancing selection (Table 4.2). This is consistent with a scenario of ongoing gene flow between species resulting in homogenisation across the genome except for regions under divergent selection (Wu 2001; Gavrilets & Vose 2005). It is possible that these narrow regions comprise

genomic islands of speciation (e.g., Turner *et al.* 2005; Nosil *et al.* 2009; Nadeau *et al.* 2012b) however further investigation would be required to ascertain their contribution to the speciation process. Surprisingly, there were equivalent levels of divergence between the allopatric comparison of *A. alcalica* / *A. grahami* as with sympatric comparisons of *A. alcalica* / *A. latilabris* and *A. ndalalani*.

The previous chapter demonstrated that there was extensive hybridisation and gene flow between sympatric *Alcolapia* species (chapter three). Incipient species may never achieve full speciation if the speciation process is reversed by interspecific hybridisation brought on by changes in the environment. However, results of the present analysis suggest that it is unlikely that *Alcolapia* are currently undergoing speciation reversal given that similar frequencies of outlier SNPs were detected in the sympatric Lake Natron species comparisons as in the allopatric *A. alcalica* / *A. grahami* comparison (Table 4.2). If speciation reversal were occurring, we would expect the sympatric species to show fewer putative outlier SNPs under diversifying selection than the allopatric species, as introgression would erode peaks of differentiation. Such an impact of hybridisation has been seen in European whitefish where a breakdown in reproductive isolating mechanisms increased gene flow between species, reducing the extent of genomic islands of differentiation, and exhibiting fewer candidate outlier loci (Vonlanthen *et al.* 2012; Hudson *et al.* 2013).

Patterns of differentiation were not as clear for intraspecific variation. Despite peaks in F_{ST} sliding window analyses for intraspecific morphs and populations (Figures 4.7 and 4.8), none of these comparisons resulted in outlier loci being detected in BayeScan analysis (Table 4.2). Furthermore, peaks of differentiation did not show evidence of elevated linkage disequilibrium (Figure 4A.3). Although the datasets for Bayescan and LD analysis were reduced from the full dataset used for genomic scans and contained fewer SNPs per genomic region (table 4.3), the filtering parameters accounted for missing data and low allelic frequency, so it is unlikely that the inclusion of the additional SNPs would have proved more informative. However, it does highlight that the peaks of differentiation are based on only a small number of SNPs so interpretation of intraspecific variation can only be tentative based on the existing data. Increased coverage of the regions of interest such as through the use of whole genome resequencing or targeted sequencing (e.g., Nadeau *et al.* 2012b; Martin *et al.* 2013) may improve the resolution of this differentiation. Whole genome sequencing would also increase number of markers available for alternative testing such as genome-wide association studies. Increased marker density would be relevant here, especially as patterns of LD were high

across all loci (linkage seen at distances of up to 200~500 Kb, chapter three) and LD can bias such association studies (Pardo-Diaz *et al.* 2015).

There were notably no clear peaks of differentiation between sexes within species, although the original sampling design was not intended to consider sex differentiation and did not include equal sample numbers of each sex, so interpretation is limited based on the small sample size available. It is also possible that any candidate sex-linked regions were not sequenced (RAD data equated to sequence coverage of ~3% of the genome), if present at all. The relative importance of genetic vs. environmental sex determination mechanisms is not known for *Alcolapia*, although the closely related *O. niloticus* has been shown to exhibit genetic determination with environmental influence (sex reversal due to elevated temperature during differentiation; Palaiokostas *et al.* 2013). Furthermore, the present analysis included only those reads mapping to the current tilapia reference genome, which is derived from an isogenic clonal female line (Brawand *et al.* 2014). Thus, if a sex-determining locus in *Alcolapia* were male-specific, it would not be present in the reference-aligned dataset.

Sliding window analyses also revealed differentiation between allopatrically separated populations. The northern *A. alcalica* clade exhibited peaks of differentiation from the southern clade, and also appeared more differentiated from other species relative to the southern clade (although most of the peaks of differentiation were shared) (Figures 4.5, 4.7). The comparison of these clades also detected outlier loci under diversifying selection in BayeScan analysis (Table 4.3). These results are congruent with the phylogenomic and population genetic results from chapter three that the northern populations appear isolated from those bordering the southern lagoon. Differentiation was also seen in the comparison of *A. grahami* from Lakes Magadi and Nakuru (although not in BayeScan analysis), suggesting that the population in Lake Nakuru experiences a different selective regime than that of Lake Magadi. Further analysis would be required to characterise the nature of the different selective pressures in different habitats. Population bottlenecks (as may have been experienced by a founder population of *A. grahami* introduced to Lake Nakuru) can have effects on the detection of selection, resulting in false positives (Foll & Gaggiotti 2008), but have been reported to result in depressed F_{ST} in instances of high heterozygosity because of model assumptions of large populations (reported for minisatellite markers) (Flint *et al.* 1999). Evidence of bottlenecks and subsequent expansion have previously been reported based on mtDNA and microsatellite data for *A. alcalica* and *A. latilabris* (Zaccara *et al.* 2014), but similar analysis revealed no evidence of population size change for *A. grahami*

Magadi populations (Kavembe *et al.* 2013). However, estimates of heterozygosity in the present analysis were not significantly different between populations of the same species (Figure 4.9), so previous demographic changes do not appear to have caused losses of heterozygosity in isolated populations.

Alcolapia relationship to *Oreochromis* outgroup

Although the present analysis did not explicitly examine *Alcolapia* phylogeny, phylogenomic incongruence was investigated based on the pattern of elevated divergence seen in sliding window D_{XY} analyses (Figure 4.2). When ML analysis was conducted on a 3Mb region in which a D_{XY} peak was seen (LG23), resolution within the *Alcolapia* radiation was poor, but patterns of diversity were clearly seen in elevated branch length relative to non-peak regions (Figures 4A.1 and 4A.2). However, increased diversity was not partitioned within species and few nodes had high BS support. These results emphasise the very recent divergence of the Alcolapia species, that even extensive alignments (~75 Kb) contain insufficient phylogenetic signal for resolution. Although, in common with the analysis of the full dataset (chapter three), these reasonably large alignments had a very low proportion of variable sites (2%). Conversely, the outgroup nodes were well supported in the linkage group analyses. Furthermore, the branch length from *Alcolapia* to *O. amphimelas* was reduced relative to *O. niloticus* in the LG23 region of D_{XY} peak, compared with the non-peak regions. This may suggest less divergence in the region than adjacent regions, and also correlates with a reduction in F_{ST} between *A. alcalica* and *O. amphimelas* seen in the same region (Figure 4.2). Although not identified by BayeScan as an outlier of balancing selection, the sliding window and phylogenomic analyses suggest that a particular region on LG23 exhibits less divergence between the soda lake species (*Alcolapia and O. amphimelas*) than the freshwater species (*O. niloticus*) relative to other regions of genome. Further analysis and identification of candidate loci within the region would be required to investigate whether the region is implicated in adaptation to soda conditions, either via introgression or convergent adaptation.

Intraspecific colour and morphology divergence

Sliding window analyses revealed distinct peaks in *A. alcalica* morphs, which were more pronounced in the mouth morphs compared to the colour morphs (Figure 4.7). Feeding specialisation leading to reproductive isolation has previously been shown

to be central to speciation in fish radiations (Nosil 2012; Bernardi 2013; Seehausen & Wagner 2014), so divergent selection acting on trophic morphology loci could lead to heterogeneous genomic differentiation, obscuring the signal of reproductive isolation in these species (although see chapter six for a full discussion of the differentiation between trophic morphs). Such localised divergence has previously been observed in recently diverged sympatric cichlid species (Franchini *et al.* 2014). We may expect to see lower genomic differentiation between intraspecific colour morphs than trophic morphs, given that colour polymorphism and pigmentation patterns have often been shown to be controlled by single genes in cichlids (Roberts *et al.* 2009; Henning *et al.* 2010; Albertson *et al.* 2014), whereas trophic morphology and jaw shape is linked to multiple genes of small effect (Albertson & Kocher 2006). Sexual selection may also be important in this system as *Alcolapia* are sexually dichromatic and polygamous. Colour pigmentation involves few genetic changes with dominant gene involvement at colour loci (discussed in Maan & Sefc 2013), so could contribute to heterogeneity across the genome. Additionally, sexual selection may permit co-existence of species occupying similar ecological niches (M'Gonigle *et al.* 2012), which may be particularly relevant here given the narrow range of trophic niches available in the present-day soda lake springs. Furthermore, it may be that the extreme conditions of the lake habitat (and subsequent requirement for extensive physiological adaptations to enable life therein) act as a constraint on the rest of the genome and limit differentiation at sites outside putative islands of divergence, providing a low tolerance to haplotype evolution beyond a narrow range. More detailed sequencing of the genome and outlier analysis, or targeted sequencing of coding regions responsible for differences in colouration and trophic morphology, would be required to test this hypothesis.

Conclusions

Evidence for narrow regions of genomic differentiation containing outlier loci is presented here in all *Alcolapia* species comparisons and intraspecific population comparisons of *A. alcalica*. Other intraspecific comparisons exhibited peaks of elevated F_{ST} in genomic scans, but no clear detection of outlier loci. The findings of the present analyses, which demonstrate low levels of genome-wide divergence alongside narrow peaks of high divergence certainly warrant further investigation in order to elucidate the processes initiating and maintaining speciation in this system. A clear future research goal would be to identify regions of the genome that are responsible for the phenotypic diversification and polymorphism observed in these

cichlid fishes, despite the very shallow divergence between species. Characterising the genomic architecture and structural variation (indels, inversions, and copy number variation) underlying regions of differentiation would also provide information on how these changes may have arisen in a recent diversification. Although the present study highlights heterogeneous genomic differentiation, more detailed analysis is required to identify the regions of high divergence and assess their impact on species differentiation.

References

Abbott R, Albach D, Ansell S *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.

Albertson RC, Kocher TD (2006) Genetic and developmental basis of cichlid trophic diversity. *Heredity*, **97**, 211–221.

Albertson RC, Powder KE, Hu Y *et al.* (2014) Genetic basis of continuous variation in the levels and modular inheritance of pigmentation in cichlid fishes. *Molecular Ecology*, **23**, 5135–5150.

Antao T, Lopes A, Lopes RJ *et al*. (2008) LOSITAN: A workbench to detect molecular adaptation based on a Fst-outlier method. *BMC Bioinformatics*, **9**, 323.

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.

Beaumont MA, Nichols RA (1996) Evaluating loci for the use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 1619–1626.

Bernardi G (2013) Speciation in Fishes. *Molecular Ecology*, **22**, 5487–5502.

Brawand D, Wagner CE, Li YI *et al.* (2014) The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, **513**, 375–381.

Charlesworth B (1998) Measures of divergence between populations and the effect of forces that reduce variability. *Molecular Biology and Evolution*, **15**, 538–543.

Clark PJ, Evans FC (1954) Distance to nearest neighbour as a measure of spatial relationships in populations. *Ecology*, **35**, 445–453.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, MA.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, **23**, 3133–3157.

Cullingham CI, Cooke JEK, Coltman DW (2014) Cross-species outlier detection reveals different evolutionary pressures between sister species. *New Phytologist*, **204**, 215–229.

De Mita S, Siol M (2012) EggLib: Processing, analysis and simulation tools for population genetics and genomics. *BMC Genetics*, **13**, 27.

Fan S, Meyer A (2014) Evolution of genomic structural variation and genomic architecture in the adaptive radiations of African cichlid fishes. *Frontiers in Genetics*, **5**, 163.

Faria R, Renaut S, Galindo J *et al.* (2014) Advances in ecological speciation: an integrative approach. *Molecular Ecology*, **23**, 513–521.

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in genetics*, **28**, 342–350.

Flint J, Bond J, Rees DC *et al.* (1999) Minisatellite mutational processes reduce F_{ST} estimates. *Human Genetics*, **105**, 567–576.

Foll M (2012) *BayeScan v2.1 User Manual*.

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, **180**, 977–993.

Franchini P, Fruciano C, Spreitzer ML *et al.* (2014) Genomic architecture of ecologically divergent body shape in a pair of sympatric crater lake cichlid fishes. *Molecular Ecology*, **23**, 1828–1845.

Gavrilets S, Vose A (2005) Dynamic patterns of adaptive radiation. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 18040–18045.

Hammond R, McCullagh P (1978) *Quantitative techniques in geography: an introduction*. Oxford University Press.

Henning F, Renz AJ, Fukamachi S, Meyer A (2010) Genetic, comparative genomic, and expression analyses of the *Mc1r* locus in the polychromatic Midas cichlid fish (Teleostei, Cichlidae *Amphilophus* sp.) species group. *Journal of Molecular Evolution*, **70**, 405–412.

Henry P, Russello MA (2013) Adaptive divergence along environmental gradients in a climate-change-sensitive mammal. *Ecology and Evolution*, **3**, 3906–3917.

Hoffman JI, Simpson F, David P *et al.* (2014) High-throughput sequencing reveals inbreeding depression in a natural population. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 3775–3780.

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.

Hudson AG, Vonlanthen P, Bezault E, Seehausen O (2013) Genomic signatures of relaxed disruptive selection associated with speciation reversal in whitefish. *BMC Evolutionary Biology*, **13**, 108.

Jones FC, Chan YF, Schmutz J *et al.* (2012) A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Current Biology*, **22**, 83–90.

Kavembe GD, Machado-Schiaffino G, Meyer A (2013) Pronounced genetic differentiation of small, isolated and fragmented tilapia populations inhabiting the Magadi Soda Lake in Kenya. *Hydrobiologia*, **739**, 55–71.

Lischer HEL, Excoffier L (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298–299.

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology*, **23**, 2178–2192.

Maan ME, Sefc KM (2013) Colour variation in cichlid fish: developmental mechanisms, selective pressures and evolutionary consequences. *Seminars in Cell and Developmental Biology*, **24**, 516–528.

Martin CH, Feinstein LC (2014) Novel trophic niches drive variable progress towards ecological speciation within an adaptive radiation of pupfishes. *Molecular Ecology*, **23**, 1846–1862.

Martin SH, Dasmahapatra KK, Nadeau NJ *et al.* (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, **23**, 1817–1828.

Martin SH, Davey JW, Jiggins CD (2014) Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular Biology and Evolution*, 1–36.

Merrill RM, Gompert Z, Dembeck LM *et al.* (2011) Mate preference across the speciation continuum in a clade of mimetic butterflies. *Evolution*, **65**, 1489–1500.

Miller JM, Malenfant RM, David P *et al.* (2013) Estimating genome-wide heterozygosity: effects of demographic history and marker type. *Heredity*, **112**, 240–247.

M'Gonigle LK, Mazzucco R, Otto SP, Dieckmann U (2012) Sexual selection enables long-term coexistence despite ecological equivalence. *Nature*, **484**, 506–509.

Nadeau NJ, Martin SH, Kozak KM *et al.* (2012a) Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular Ecology*, **22**, 814–826.

Nadeau NJ, Whibley A, Jones RT *et al.* (2012b) Genomic islands of divergence in hybridizing Heliconius butterflies identified by large-scale targeted sequencing. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, **367**, 343–353.

Narum SR, Hess JE (2011) Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources*, **11**, 184–194.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences of the United States of America*, **76**, 5269–5273.

The Marie Curie Speciation Network. (2012) What do we need to know about speciation? *Trends in Ecology & Evolution*, **27**, 27–39.

Noor MAF, Bennett SM (2009) Islands of speciation or mirages in the desert? Examining the role of resricted recombination in maintaining species. *Heredity*, **103**, 439–444.

Nosil P (2012) *Ecological speciation*. Oxford University Press, Oxford.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Palaiokostas C, Bekaert MEL, Khan MGQ *et al.* (2013) Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) using RAD sequencing. *PLoS ONE*, **8**, e68389.

Pardo-Diaz C, Salazar C, Jiggins CD (2015) Towards the identification of the loci of adaptive evolution. *Methods in Ecology and Evolution*, **6**, 445–464.

Perez-Figueroa A, Garcia-Pereira MJ, Saura M, Rolan-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, **23**, 2267–2276.

Puebla O, Bermingham E, McMillan WO (2014) Genomic atolls of differentiation in coral reef fishes (*Hypoplectrus* spp., *Serranidae*). *Molecular Ecology*, **23**, 5291–5303.

Renaut S, Maillet N, Normandeau E *et al.* (2011) Genome-wide patterns of divergence during speciation: the lake whitefish case study. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, **367**, 354–363.

Roberts RB, Ser JR, Kocher TD (2009) Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. *Science,* **326**, 998–1001.

Seeb LW, Waples RK, Limborg MT *et al.* (2014) Parallel signatures of selection in temporally isolated lineages of pink salmon. *Molecular Ecology*, **23**, 2473–2485.

Seehausen O, Wagner CE (2014) Speciation in freshwater fishes. *Annual Review of Ecology, Evolution, and Systematics*, **45**, 621–651.

Seehausen O, Butlin RK, Keller I *et al.* (2014) Genomics and the origin of species. *Nature Reviews Genetics*, **15**, 176–192.

Smadja CM, Butlin RK (2011) A framework for comparing processes of speciation in the presence of gene flow. *Molecular Ecology*, **20**, 5123–5140.

Stinchcombe JR, Hoekstra HE (2007) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, **100**, 158–170.

R Core Development Team (2014) R: A language and environment for statistical computing. *R Foundation for Statistical Computing*.

Turner TL, Hahn MW (2010) Genomic islands *of* speciation or genomic islands *and* speciation? *Molecular Ecology*, **19**, 848–850.

Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, **3**, e285.

Via S (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 9939–9946.

Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, **17**, 4334–4345.

Vonlanthen P, Bittner D, Hudson AG *et al.* (2012) Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature*, **482**, 357–362.

Wright S (1943) Isolation by distance. *Genetics*, **28**, 114–138.

Wu C-I (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

Zaccara S, Crosa G, Vanetti I *et al.* (2014) Genetic and morphological analyses indicate high population mixing in the endangered cichlid *Alcolapia* flock of East Africa. *Conservation Genetics*, **15**, 429–440.

Appendix 4

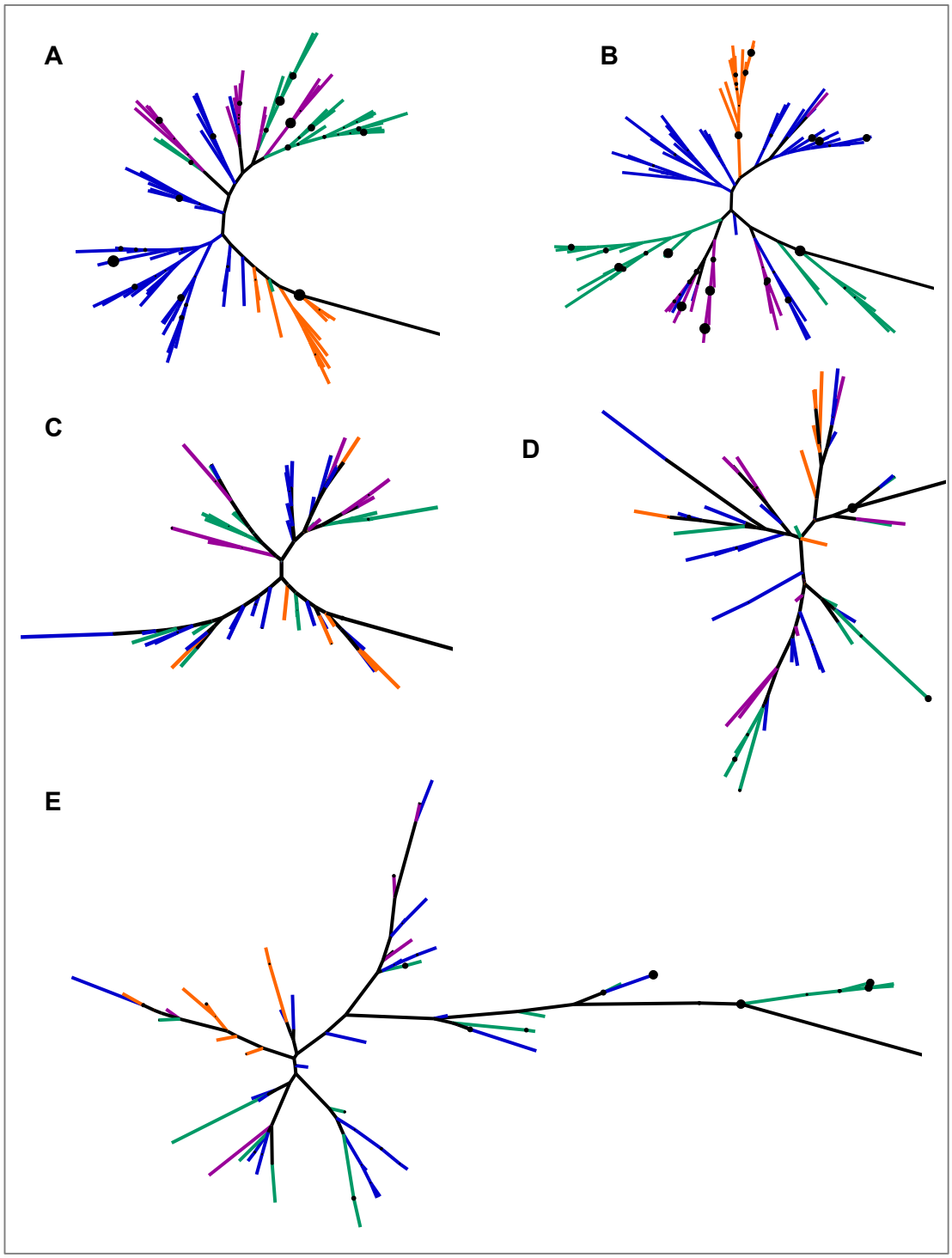


Figure 4A.1 Maximum likelihood trees of linkage groups.

Branches are coloured by species, black circles indicate BS support, scaled by value. Outgroup branches have been truncated for clarity. A) ML analysis of LG22, 10 BS; B) LG23; 10 BS; C) non-peak region of LG23 (122 Kb); 100 BS; D) non-peak region of LG23 (79 Kb), 100 BS; E) D_{XY} peak region of LG23 (74 Kb); 100 BS.

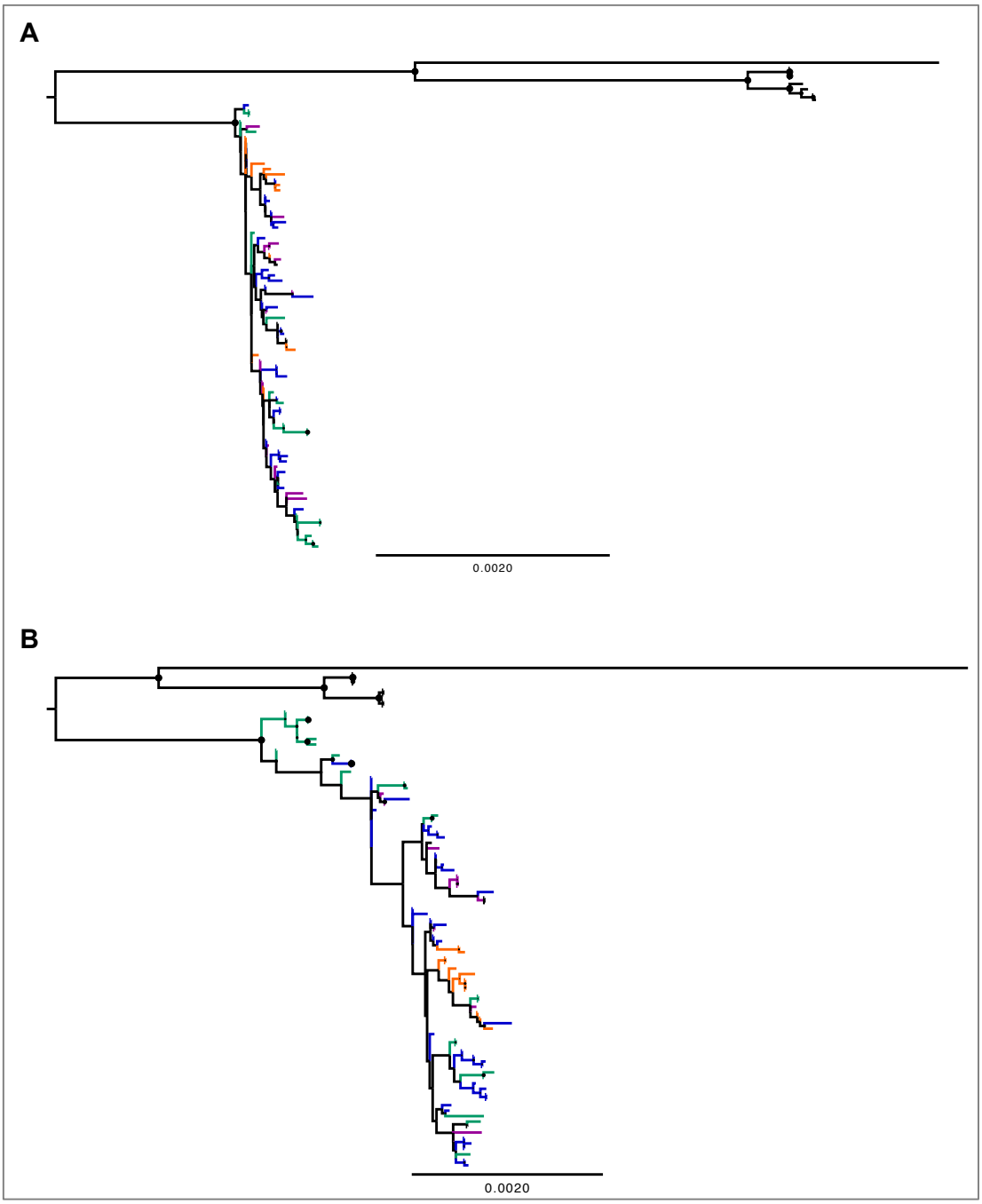


Figure 4A.2. Maximum likelihood trees of linkage group regions.

Analysis and colours as for figure 4A.1, but given in cladogram format for clarity and indication of branch length. A) non-peak region (79 Kb); B) D_{XY} peak region (74 Kb). Scale bars for both figures are 0.002.

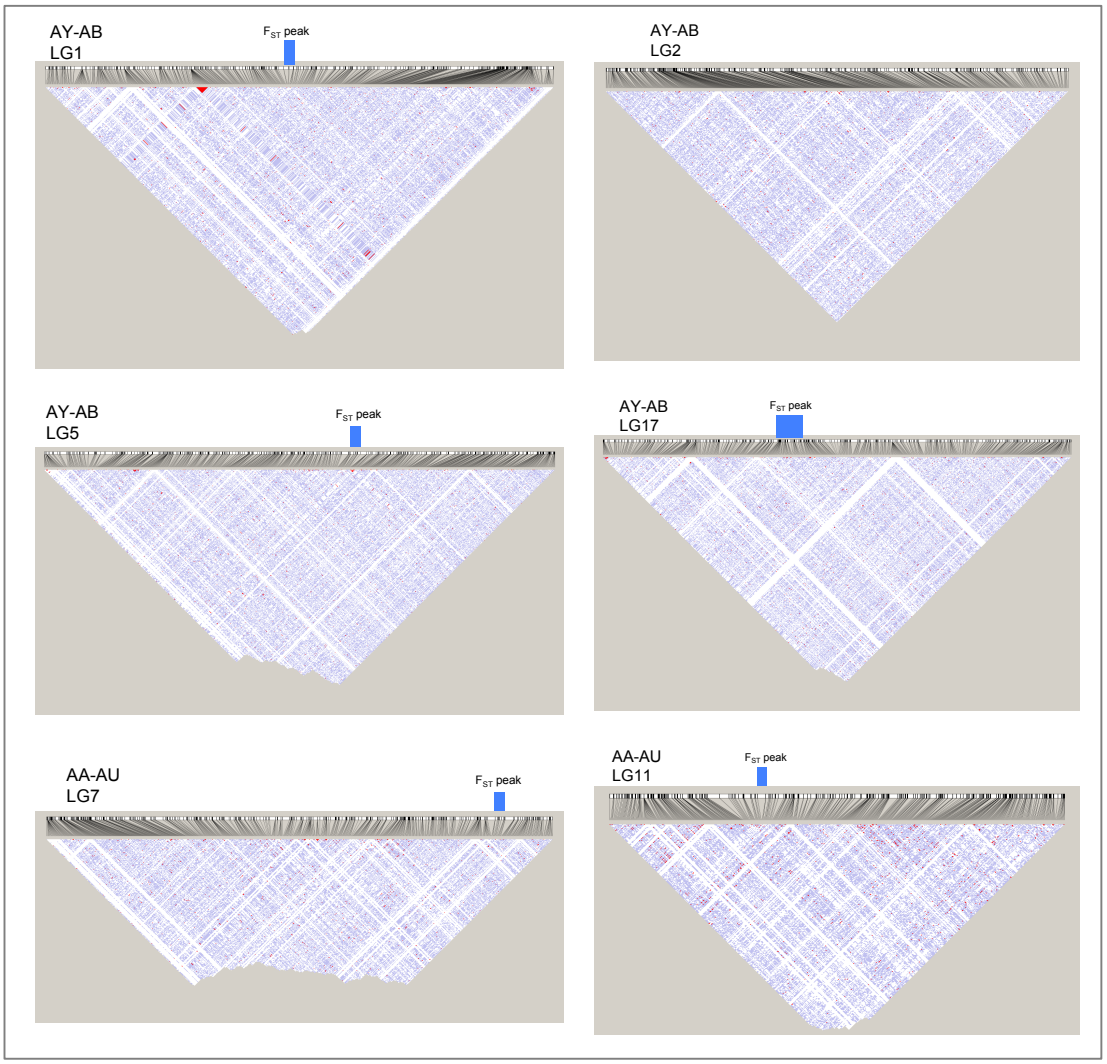


Figure 4A.3. Linkage heatmaps for linkage groups containing F_{ST} peaks.

Intraspecific comparison and linkage group is indicated in the top left-hand corner of each plot. F_{ST} peak locations are indicated by a blue bar. Linkage group 2 of the AY-AB comparisons is included as a non-peak LG comparison. Linkage blocks are coloured by D', with the following colour coding: white: LOD<2, D'<1; blue: LOD<2, D'=1; red: LOD>2.

Chapter five

Ecological and morphological divergence of East African soda lake cichlids

Abstract

Ecomorphological differentiation is a key feature of adaptive radiations, with a general trend for specialisation and niche expansion following divergence. Chapter five examines the ecological and morphological divergence of the *Alcolapia* species flock. Stable isotope data (as a proxy for ecological niche) and morphometric data are analysed in a phylogenetic framework, and contrasted with the genomic differentiation of the species flock. The analyses are also conducted at an intraspecific population level to consider ecomorphological differentiation over geographic scales. Species divergence is observed in both ecology and morphology, supporting the importance of ecological speciation within the radiation. Shallow genomic differentiation alongside large-scale ecomorphological divergence indicates rapid ecological differentiation, and three distinct regime shifts are proposed.

Introduction

The generation of phenotypic diversity is of considerable interest in evolutionary biology. Understanding how phenotypic traits evolve in closely related species may clarify the interaction of environment and morphology, particularly in cases where species experience environmental gradients across their range. The link between morphology and habitat is most commonly reported in cases of adaptive radiation, where rapidly diversifying lineages adapt to different unexploited ecological niches, and has been recorded across varied taxa, including anole lizards (Losos *et al.* 1994), three-spined stickleback fishes limnetic and benthic species pairs (Schluter & McPhail 1992), Darwin's finches (Grant & Grant 2008), European whitefish (Praebel *et al.* 2013), and African cichlid fishes (Fryer & Iles 1972). Furthermore, the

correlation of environment and morphology is clearly seen in ecomorphs of the same species adapted to different habitats, as found in wall lizards (Kaliontzopoulou *et al.* 2010), ecological races of anoles lizards (Malhotra & Thorpe 1991), benthic and limnetic morphs of Arctic charr (Snorrason & Skulason 2004), and Neotropical cichlid species (Elmer *et al.* 2010). Beyond correlation of habitat, testing if selection operates on morphology differentially in varying environments is less well studied, although attempts have been made to unify the factors of morphology, performance and fitness in the ecomorphological paradigm (Arnold 1983). However, a more recent review of studies quantifying selection on performance found no evidence that selection was stronger on performance traits than on morphological traits across a diverse range of animal taxa (Irschick *et al.* 2008). We may expect populations occurring in sympatry to exploit more differentiated niches, as competition is known to promote divergence in adaptive radiation (Schluter 1994), and once specialisation has occurred, lineages do not tend to revert to a more generalised form even if competition is reduced or conditions change (Losos *et al.* 1994).

The cichlid fishes of East Africa are well known for their diverse range of trophic adaptations to varied ecological niches, encompassing several independent adaptive radiations, and with parallel morphologies often seen between radiations and within lineages (Fryer & Iles 1972; Rüber *et al.* 1999; Albertson & Kocher 2006). Trophic diversity includes not only specialisation in resource utilisation (such as herbivory, carnivory, or piscivory), but also on food size and habitat type, e.g., algivorous species may be further segregated by targeted resource size/depth and substrate type/slope (reviewed in Burress 2014; Seehausen & Wagner 2014). Such extensive levels of resource partitioning, along with colour differentiation is posited to enable rapid speciation in cichlid fish with only subtle differences in feeding behaviour or morphology (discussed in Schluter 2000). Adaptive radiation may be facilitated and accelerated by a versatile morphology that can rapidly be modified to create varied phenotypes adapted to different environments (Schluter 2000), and such versatility may be increased by integration of separate components that can be modified independently (Vermeij 1973). Such versatility in cichlid fishes is provided by the presence of pharyngeal jaws that free oral jaws for modification to food foraging rather than processing (Liem 1973); suturing of the lower pharyngeal jaw that increases force exertion and allows adaptation to durable foods (Hulsey 2006); and decoupling of the upper and lower jaw that allows independent movement (Galis & Drucker 1996).

Cichlid jaws show rapid morphological adaptation and diversification based on diet, modified over the course of only a few generations after diet changes in the case of oral jaws (van Rijssel *et al.* 2014) and pharyngeal jaws that appear phenotypically plastic and largely influenced by diet, changing across the lifetime of an individual (Muschick *et al.* 2011; Gunter *et al.* 2013). As such, examination of jaw morphology may provide an indication of current or very recent ecological trophic niche, and provide a basis for examining adaptive divergence based on food type and foraging behaviour (e.g., Hulsey *et al.* 2008; Muschick *et al.* 2012; Theis *et al.* 2014).

Fish body shape not only varies with evolutionary relationships, but is also influenced by environmental conditions, in particular salinity, flow and water chemistry (Gomes & Monteiro 2008; Firmat *et al.* 2012). Body shape is expected to be of particular ecological relevance in fish, and morphometric analysis of body shape has successfully been used to test occurrences of ecological speciation in recent cichlid radiations (Rüber & Adams 2001; Klingenberg *et al.* 2003; Elmer *et al.* 2010; Colombo *et al.* 2012; Martin 2013). Geometric morphometric analysis condenses shape information about a configuration of landmarks (removing variation in size, position and orientation), establishing a one-to-one correspondence between individual specimens, and allowing shape data to be analysed in combination with molecular and ecological data (Klingenberg *et al.* 2011). See chapter two for a more detailed discussion of morphometric methods.

As well as studying evolution and diversification in fishes through adaptive and plastic changes in morphology, information on the ecological segregation of co-occurring species can inform the nature and extent of niche partitioning. Stable isotope analysis (SIA) is a useful tool for examining trophic niche based on specific isotopic signatures ($\delta^{13}C$, $\delta^{15}N$) acquired by consumers from different food sources to determine whether sympatric species are separated by diet (Genner *et al.* 1999; Post 2002), and whether community-wide trophic metrics differ between sites (Layman *et al.* 2007). The isotope ratios provide a useful proxy for identifying interspecific feeding differences over the long term by analysing element assimilation into tissue (versus short-term measures of feeding, such as gut contents analysis; e.g., Marijnissen *et al.* 2008). Measuring the elemental isotopic ratios in organic samples allows trophic niche to be differentiated as nitrogen is enriched through each trophic level (so providing information on trophic position of consumers) and carbon levels remain unchanged through the food chain, but are dependent on basal resources. Such characterisation allows inferences to be made regarding the role of ecological speciation in adaptive radiation, as to whether

species occurring sympatrically exhibit different diet specialisations from one another.

The species designations of the *Alcolapia* species flock are based on morphology and male colour (see Table 1.3 in chapter one). As these descriptions are based on meristic counts and measurements, a geometric morphometric approach to consider morphological differentiation will be useful in this system to consider overall body shape divergence between species, and in providing continuous variables of morphological distance for comparison to other datasets. Previous morphometric analysis of Lake Natron species using discrete measurements has shown significant differentiation between the species living in sympatry, which was maintained between different populations (Zaccara *et al.* 2014). To my knowledge, no morphometric comparison of *A. grahami* and Lake Natron species has been made, although a previous study found that *A. alcalica* displayed an intermediate relative gut length compared to that of Lake Magadi *A. grahami* populations (Wilson *et al.* 2004). Thus, this chapter aims to address the morphometric component of shape variation using geometric morphometrics (GMM) across the entire species flock for the first time. Previous gut content analysis in *A. grahami* has identified a diet comprising ~90% algal matter (including cyanobacteria) and <10% invertebrates (copepods and dipterous larvae) (Coe 1966), although nothing is known about the diet of Lake Natron species. Investigation of trophic niche using stable isotope analysis, relative gut length, and stomach contents, is pertinent for this system, considering that the species are largely differentiated based on trophic morphology. Thus, the use of GMM in combination with ecological (stable isotope analysis, inferences from pharyngeal jaw shape) and genomic data (chapter three), will add to previous work on this system by allowing reconstruction of the process of morphological diversification (using ancestral state reconstruction), and examination of the extent of correlation of phenotype with environment.

Lake Natron contains populations where all species occur in sympatry and monospecific sites where only *A. alcalica* is found. Genomic analysis indicates that isolated populations of *A. alcalica* (outside the southern lagoons) are more genetically differentiated and experience a lower degree of interspecific gene flow than those populations occurring in sympatry with *A. latilabris* and *A. ndalalani* (Ford *et al.* 2015, chapter three). Analysis of multiple Lake Natron populations occurring both in sympatry and allopatry, as well as single species populations of *A. grahami* in Lake Magadi, may allow insight into the selective pressures affecting these fish when faced with interspecific competition and how niche partitioning may change

with environment. Furthermore, previous field observations suggested that colour morphs (light/dark) of *A. alcalica* exhibited different behaviour and inhabited different microhabitats (muddy/clear water) (Seegers & Tichy 1999).

Aims

The integration of multi-disciplinary methods using ecology, morphology and genetics to examine recently diverged species has been used to disentangle the processes underlying the generation and maintenance of diversity (Gavrilets & Losos 2009; Elmer *et al.* 2010; Muschick *et al.* 2012; Martin 2013; Muschick *et al.* 2014). To investigate diversification in the unique *Alcolapia* cichlid radiation a combination of stable isotope analysis (as a proxy for niche space) and geometric morphometrics is used here to analyse body and lower pharyngeal jaw shape, and combined with a densely sampled genome-wide SNP dataset. Ecological and morphological segregation is investigated to test the prediction that ecological speciation has been an important driver of adaptive diversification in the soda lake cichlids. Ecomorphology of the flock is considered within a phylogenetic framework to test i) the degree of niche partitioning between sympatrically occurring species; ii) the degree to which competition drives feeding specialisation, by comparing trophic specialisation between sympatric and monospecific sites; iii) whether morphological adaptation is correlated with ecological niche utilisation; iv) the correlation of ecological and morphological differentiation with phylogeny.

Methods

Sampling

Samples were collected in 2012 using hand, cast or seine nets dependent on substrate type and water depth (Permit numbers: NCST/RCD/126/012/29 and 012-25-NA-2011-182). Fish were euthanised using tricaine methanesulfonate (MS222) and preserved as voucher specimens in 80% ethanol, with genetic samples (fin clips) stored in 95% ethanol for later analysis. Sampling locations are displayed on the map in Figure 5.1.

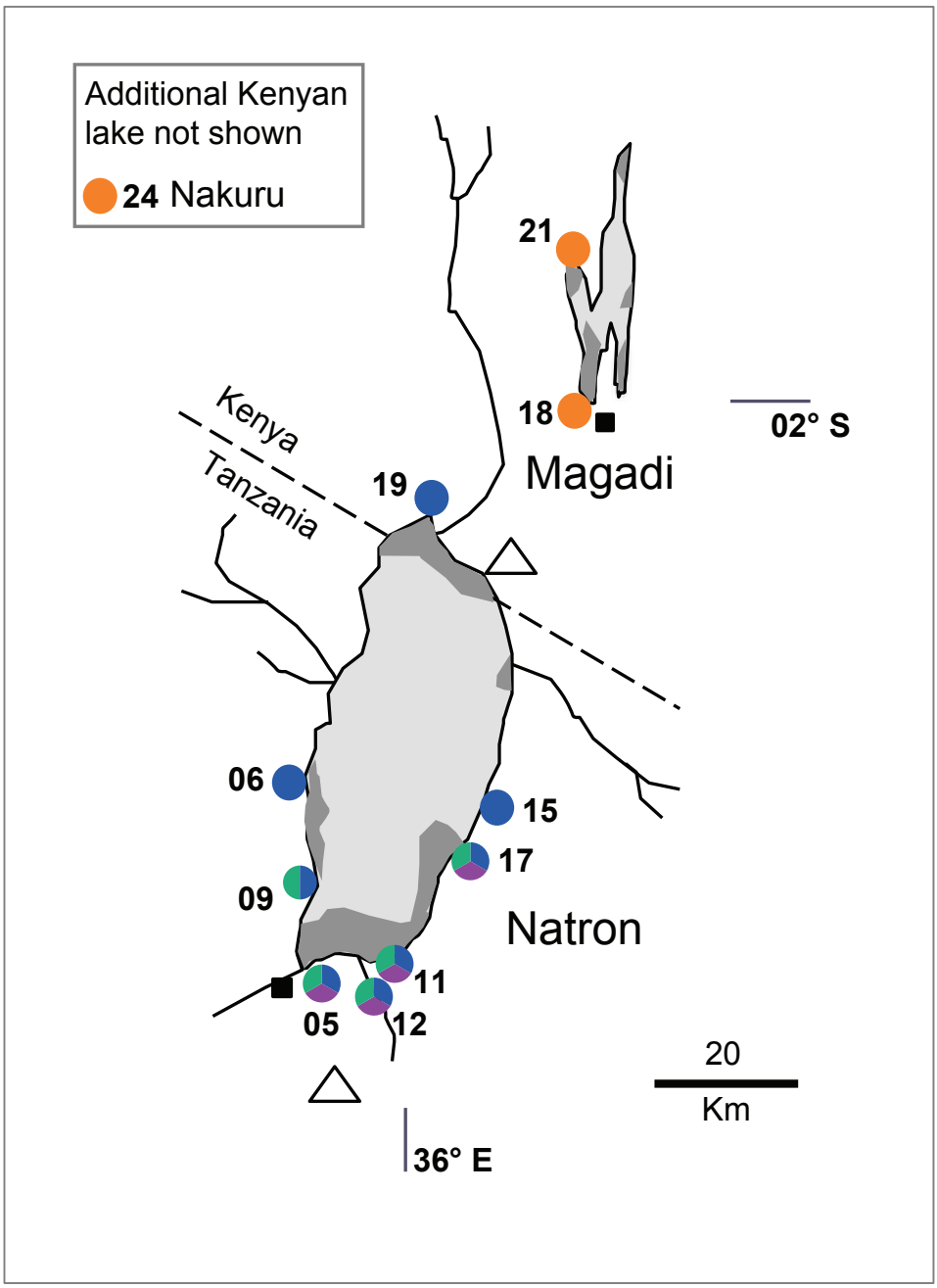


Figure 5.1. Sampling locations for the present study. Site markers are coloured by species present (not proportional to species abundance). Blue: *A. alcalica*; Green: *A. latilabris*; Purple: *A. ndalalani*; Orange: *A. grahami*. Black squares: type localities for *Alcolapia* species; open triangles: volcanoes. Lake basins are outlined in black, with light grey shading representing trona crust, and dark grey indicating areas of open water (lagoons). Lake Natron has several perennial inflowing rivers and streams (black lines).

Sample sizes and data subsets

All datasets were analysed separately before being integrated with other datasets for comparison. As such, the full sample size available from each methodology was included for preliminary analysis of each dataset: RAD (n=91); stable isotope analysis (n=458); body shape geometric morphometrics (n=735). Consequently, all tests for potential biases or preservation effects described below were undertaken on the full datasets. Thereafter, datasets were integrated for comparison and only subsets of the stable isotope and morphometrics were retained in order to restrict inclusion to sampling sites for which all data types were available. The results presented in this chapter are therefore from the analysis of data subsets, although preliminary processing tests were not repeated once data was subdivided. The full datasets are analysed in further detail in chapter six. Total specimen numbers by analysis, along with site information and GPS coordinates are presented in chapter six and Appendix seven.

Table 5.1. Sampling locations and specimen numbers by analysis.

LPJ: Lower pharyngeal jaw; RAD: Restriction site-associated DNA (genomic data); SIA: Stable isotope analysis.

Sampling site GPS coordinates and specimen IDs for each analysis are given in Appendix 7.

Lake	Site	Species	RAD	SIA	Stomach contents	GMM		Covariation		
						Body	LPJ	SIA/Body	SIA/RAD	RAD/Body
Natron	005	*A. alcalica*	4	13	10	13	10	12	4	4
		A. latilabris	4	15	13	31	10	13	4	4
		A. ndalalani	4	12	11	32	10	10	3	4
	006	*A. alcalica*	4	15	-	24	-	13	3	4
	009	*A. alcalica*	4	16	-	14	-	11	4	3
		A. latilabris	4	15	-	16	-	11	4	4
	011	*A. alcalica*	8	29	10	73	-	26	8	6
		A. latilabris	4	19	12	66	-	18	4	4
		A. ndalalani	4	30	12	63	-	27	4	4
	012	*A. alcalica*	2	15	12	16	12	10	2	1
		A. latilabris	3	15	10	22	15	13	3	3
		A. ndalalani	3	15	11	22	13	15	3	3
	015	*A. alcalica*	4	15	12	27	11	15	4	4
		A. alcalica (upturned)	4	16	-	30	10	15	4	3
	017	*A. alcalica*	4	4	-	0	-	-	-	-
		A. latilabris	4	16	-	31	-	11	4	3
		A. ndalalani	4	14	-	17	-	15	4	4
	019	*A. alcalica*	4	13	-	14	8	14	4	3
Magadi	018	*A. grahami*	4	16	11	18	-	15	4	4
	021	*A. grahami*	4	15	10	11	9	10	4	2
Nakuru	024	*A. grahami*	4	15	-	27	-	12	3	2
Total			**84**	**360**	**134**	**567**	**108**	**286**	**77**	**69**

Stable isotope analysis

Sample preparation and processing

Stable isotope ratios of carbon (^{13}C) and nitrogen (^{15}N) were analysed using continuous flow isotope ratio mass spectrometry (CF-IRMS) for n≈15 for each cichlid species per site (see Table 5.1), along with baseline samples of algae and invertebrates where available. Stable isotope ratios are given using the δ notation expressed in units per mille as follows:

$$\delta\ (‰) = [(R\ sample\ /R\ standard)–1] \times 1000, \text{ where } R = {}^{13}C/{}^{12}C \text{ or } {}^{15}N/{}^{14}N.$$

White muscle tissue, removed from the right-hand flank, was used for SIA. Tissue samples for a total of 15 individuals across multiple sites were dissected on-site in the field and air-dried to provide control measures. The remaining sample tissues were removed from ethanol-preserved fish (including those from which control samples had been taken) following fieldwork. Samples were dried for 24–48 h at 60°C and ground to a homogenous powder using a pestle and mortar. The ground samples were weighed into tin capsules at weights of 0.7±0.1 mg for animal tissues, and 1.2±0.1 mg for plant matter (some algae samples required larger weight for accurate analysis – where this was the case, mass was calculated based on the $\delta^{13}C$ values from the initial run, and the sample rerun with a sample of the required weight). The weighed samples were analysed using a Delta V Plus (Thermo Scientific) Mass Spectrometer with an ECS 4010 elemental analyser (Costech instruments) at the Scottish Universities Environmental Research Centre (SUERC) Life Sciences Mass Spectrometry Facility, East Kilbride, UK. Gelatine, alanine, and glycine were used as laboratory standards (drift standards) and tryptophan or glutamic acid (USGS 40) were used as elemental standards for C and N_2. The samples were analysed in 10 separate runs, and within-run standard deviation for both $\delta^{15}N$ and $\delta^{13}C$ of the standards was ≤0.20‰ for all runs.

Lipid correction

Lipids are known to be depleted in ^{13}C relative to protein (DeNiro & Epstein 1977) owing to the different chemical pathways in synthesis, and so a bulk $\delta^{13}C$ value from a sample containing both lipid and protein will not accurately reflect diet. The use of non-fatty tissue is thought to obviate this potential discrepancy, and therefore white muscle tissue is considered the best tissue to use for SIA based on its low lipid content and smallest variation between samples of all tissues (Pinnegar & Polunin

1999). However, muscle tissue may not be completely devoid of lipid, and therefore could still be affected by depletion of ^{13}C. To check for ^{13}C depletion in the present analysis, ratios of carbon and nitrogen were compared, as a proxy for lipid concentration, using C:N ratio by weight (rather than the molecular ratio, although both values have previously given identical results in arithmetic corrections; Sweeting *et al.* 2006). C:N ratio varied both between species (ANOVA: $F=15.71$, $P<0.001$) and within species between sites (*A. alcalica* - ANOVA: $F=14.25$, $P<0.001$; *A. grahami* - ANOVA: $F=34.29$, $P<0.001$; *A. latilabris* - ANOVA: $F=4.86$, $P<0.001$; *A. ndalalani* - ANOVA: $F=3.63$, $P<0.01$). As chemical lipid extraction significantly increases time and cost of analysis (since separate samples must be used for $\delta^{13}C$ and $\delta^{15}N$ analysis), lipid normalisation was conducted using C:N ratios with an arithmetic correction technique. All fish sample ^{13}C values were lipid corrected based on C:N ratio, using equations 1 and 5 (and estimated parameters) from Kiljunen *et al.* (2006), and these corrected values were used for all subsequent analysis. The $\delta^{13}C$ values for invertebrate and algae samples were not corrected.

Tissue preservation effect

Sample collection locality and remoteness meant that optimal preservation methods (freezing or drying) could not be used for all samples, and so samples were preserved in 80% ethanol. As chemical preservation may affect stable isotope values (e.g., Kelly *et al.* 2006; Correa 2012), within-sample comparisons were performed between the air-dried and ethanol-preserved tissues of the 15 control samples. Significant $\delta^{13}C$ enrichment was seen in ethanol-preserved samples for the raw carbon values (mean enrichment of 0.83‰; Wilcoxon signed rank test: $V=47$, $P<0.001$; Paired t test: $t=-4.9787$, $P<0.001$), however the effect was not significant in the C:N lipid-normalised $\delta^{13}C$ values (Wilcoxon signed rank test: $V=0$, $P=0.761$; Paired t test: $t=0.9625$, $P=0.353$), which is likely due to the fact that ethanol preservation is associated with lipid loss and leaching (Vizza *et al.* 2013). Therefore, no correction was applied for ethanol preservation, as the lipid-corrected values were used for all further analyses.

Body size effect

Body size may affect trophic position and niche space due to ontogenetic shifts in diet at different life stages (Hjelm *et al.* 2001; Post 2003) and differently sized fish may inhabit different extremes of the local environment e.g., water depth (Correa *et al.* 2012). However, it is unlikely that water depth would have an effect in this system

given the extremely shallow lake and spring water, with all sites in the present study <1.5m deep. Specimens included in the present analysis were all adult fish to reduce impact of any ontogenetic change in diet. However, as there remained a considerable size range even in the adult fish (standard length, SL: 26–104 mm) stable isotope values were tested for an effect of body size. Comparisons were made for the total dataset and species- and site-specific subsets of the data. There was no consistent effect of body size (standard length; SL) on $\delta^{13}C$ across 42 comparisons (15 showed a significant effect of body size (Pearson product moment correlation $P<0.05$), of which 10 exhibited positive correlations and 5 showed negative correlations), or $\delta^{15}N$ (for 42 comparisons, 14 showed a significant effect of body size, of which 3 exhibited positive correlations and 11 showed negative correlations). Significant comparisons for $\delta^{15}N$ were different subsets than those that were significant for $\delta^{13}C$. As there was no consistent effect of body size, no correction was applied to account for differences in body size between samples.

Statistical analysis

All analyses were conducted on the lipid-corrected $\delta^{13}C$ and raw $\delta^{15}N$ values as explained above. Food partitioning between species at sympatric locations was tested using ANOVA across isotope means between species in each site. Total isotopic niche space was analysed by standard ellipse area adjusted for small sample size (SEAc) (Jackson *et al.* 2011) implemented in the R package SIAR (Stable Isotope Analysis in R; Parnell *et al.* 2010). Pairwise dietary distances were considered in a matrix of isotopic distances between individuals at each site, calculated by treating the $\delta^{13}C$ and $\delta^{15}N$ values as Cartesian coordinates using the dist function in R 2.15.1.

Absolute isotopic values are not directly comparable between different sites, as differing system inputs mean that individual sites can vary in baseline isotopic vales of nitrogen and carbon. As such, baseline levels may be established using primary producer and prey item samples from each site (Post 2002), and correcting the values for fish samples correspondingly. However, in the current study insufficient baseline data were available to baseline-correct samples at all sites (due to a paucity of invertebrate and algal samples collected at fieldwork sites) and so cross-site comparisons were not performed here.

Furthermore, tissue nitrogen isotope ratios have been shown to vary as an effect of microhabitat (ambient nitrogen levels) and nitrogen metabolism (Moeri *et al.* 2003), which would be a consideration for *Alcolapia* in being ureotelic rather than

ammoniotelic, but is not expected to affect the current study as all species are believed to exhibit the same nitrogen metabolism mechanism (although shown experimentally only in *A. grahami* and *A. alcalica*; (Wilson *et al.* 2004).

Gut length and stomach contents

Data collection

Voucher specimens for morphological analysis were stored in 70-80% ethanol. Standard length of preserved specimens was measured using digital callipers. Intestines and stomachs were removed via ventral incision in the body wall, and intestines were uncoiled and measured from the anus to the stomach using a ruler. Specimens for which intestines stretched or disintegrated during uncoiling were excluded. Stomachs were dissected under a binocular microscope (Leica) and contents separated into the following categories: algae and cyanobacteria; cellulose and plant material (including seeds); small arthropods (insects and zooplankton); hard-shelled invertebrates; fish fry and eggs; fish remains and scales; grit and sand. As no hard-shelled invertebrates were found in any of the fish, this category was discarded from further analysis. Proportion by volume (percentage) was estimated against volume for each category per individual. Specimens with entirely empty stomachs were not included in the analysis, although intestine length was still measured, resulting in slight differences in the number of specimens included in stomach contents or gut length analysis.

Data analysis

Proportion of stomach contents was analysed using Schoener's index of dietary overlap (Schoener 1970), calculated in the FSA package in R 3.12 (Ogle 2015). Intestine length and body (standard) length values were $\log_{10}$-corrected to homogenise variance. For the pairwise comparisons of Schoener's index, a value of >0.6 was considered to represent substantial ('relevant') biological overlap (Wallace 1981).

As there is an allometric relationship of intestine length with body length in fish (Kramer & Bryant 1995), ideally body mass would also be used to correct for allometric effects (e.g., Wagner *et al.* 2009; Davis *et al.* 2013). However, as weight was not recorded in the field and tissue samples had already been taken from preserved specimens prior to gut and stomach analysis, body mass data were not available for the present study. Instead, relative gut length ($\log_{10}$gut length /

log_{10}standard length) was used to assess differences between species and populations. Group means were tested for significant differences using ANOVA. Given the close relationships of the focal species, data were not phylogenetically corrected prior to analysis.

Geometric morphometrics – Body shape

Data collection

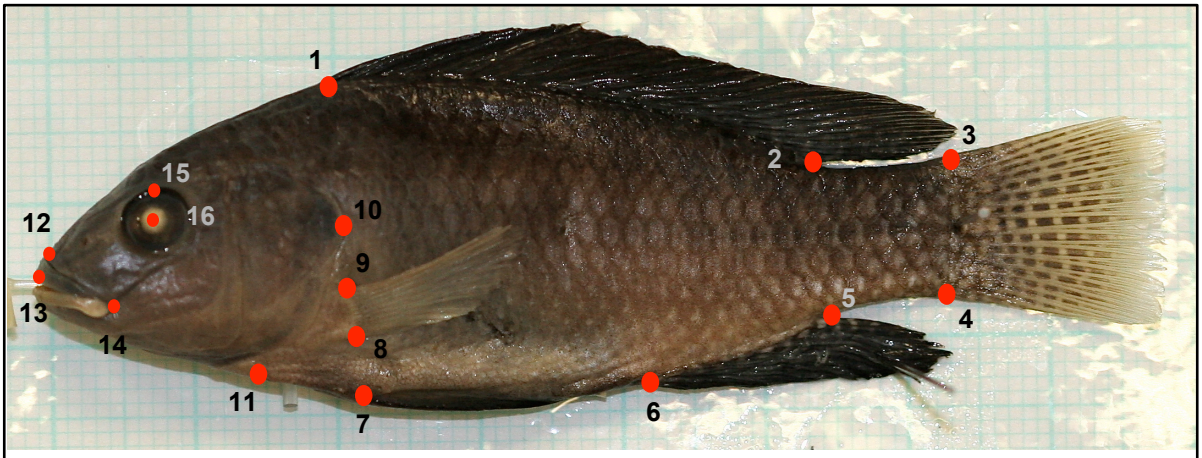
Morphological differentiation was analysed using geometric morphometric analysis of 2D digital photos. Photographs were taken of the left-hand side of ethanol-preserved specimens from a set distance of 0.5m using a tripod and Canon EOS 2OD DS126061 camera with Macro lens EF 100m 1:2.8 USM. Specimens were pinned if necessary to clarify relevant anatomical features and minimise any warping from preservation effects.

External sexing was only possible for dominant (displaying) males based on colour, and for brooding females based on fry-carrying in mouth. As such, it was not possible to sex the majority of adult fish. Therefore, individuals were not analysed based on sex for morphological analysis, and all individuals were analysed by species or population only. Individuals included for analysis were selected dependent on preservation quality alone, with all voucher specimens in each population included if suitable for morphometric analysis.

Data analysis

Digital images were processed in tpsUtil v 1.58 and landmarks were digitised using tpsDig2 v 2.17 (Rohlf 2013). Twenty-five homologous landmarks were tested for reproducibility in a pilot study across two of the study sites (site 005 and site 009), from which the final set of 16 landmarks for analysis were selected (Figure 5.2). The final landmark set was tested by replicating data collection twice across one site (site 005) in a blind test – values did not differ significantly between rounds of data collection (Procrustes ANOVA: F=0.85; P=0.8266); thereafter data collection consisted of just one round of landmark placement.

To consider species-level relationships of the *Alcolapia* flock, the morphological dataset was analysed including only populations for which genomic data were available (allowing phylogenomic comparison and assessment of phylogenetic signal). Additional populations (for which genomic data were not available) are included in a larger analysis to consider the morphological intraspecific variation across a greater number of populations in chapter six.



1. Anterior insertion of dorsal fin
2. Posterior insertion of dorsal fin
3. Dorsal insertion of caudal fin
4. Ventral base of caudal fin
5. Posterior insertion of anal fin
6. Anterior insertion of anal fin
7. Anterior insertion of pelvic fin
8. Lower insertion of pectoral fin
9. Upper insertion of pectoral fin
10. Posterior extremity of operculum
11. Anteroventral tip of the pectoral girdle
12. Anterior tip of maxilla
13. Lip juncture (anterior snout tip)
14. Posterior point of mouth cleavage
15. Dorsal margin of the eye
16. Centre of the orbit

Figure 5.2. Body shape landmarks for geometric morphometric analysis.

Morphometric analysis was conducted in MorphoJ v 1.05f (Klingenberg 2011), using a Procrustes superimposition (Rohlf & Slice 1990) to remove size and orientation differences. Data was checked for outliers in MorphoJ (no outliers required removal). To correct for any impact of allometry, a regression was performed of Procrustes coordinates against centroid size (where size accounted for 5.6% of total dataset variation; 10,000-round permutation test against null hypothesis of independence: $P<0.0001$) and the resulting residuals of this regression were used for all downstream analyses.

Comparisons were conducted on the entire dataset and subdivisions by site and species. The *A. alcalica* colour morphs (blue and yellow) from site 11 were compared separately, but as these showed no significant morphometric differentiation from each other (Discriminant function analysis – Procrustes distance: 0.0271 (P=0.4870); Mahalanobis distance: 1.2630; T-square: 2.1270 (P=0.3320)), were recoded and included alongside the remaining *A. alcalica* specimens for all further analyses.

Principal component analysis (PCA) and canonical variate analysis (CVA) were conducted for pairwise comparisons between species and sites. Scatter plots of PCA and CVA were produced with confidence ellipses of the mean at the 90% level. Shape changes were visualised using the thin-plate spine, and all diagrams

produced to the default scale factor of 1.0, or to the maximum scale of the specific axis of variation as applicable. Interspecimen pairwise Mahalanobis and Euclidean distances, group NPMANOVA (non-parametric multivariate analysis of variance) tests, and hierarchical clustering using Ward and neighbour-joining algorithms were calculated in PAST v 2.17c (Hammer *et al.* 2001) using the Procrustes-fitted regression residuals, and where relevant all tests were performed with 10,000 permutations for p-values.

The correction of morphometric data for phylogeny prior to analysis has previously been shown to obscure evolutionary relationships among traits when considered in an adaptive context (Polly *et al.* 2013). However, as such correction is commonly employed when comparing species-level taxa, data were checked for phylogenetic signal using the populations for which genomic data was available. The correlation was significant in permutation testing in MorphoJ (see Results), suggesting a significant effect of phylogeny on the dataset. Therefore, a phylogenetic principal component analysis (pPCA) was conducted to assess the impact of the phylogenetic signal, and all analyses were conducted in Mathematica 10.0.2.0 (Wolfram Research, Inc.) using packages Geometrics Morphometrics 11.0 and Phylogenetics 3.0 (Polly 2014). As the pPCA produced no discernible difference in the PCA (see Results), no phylogenetic correction was applied to the dataset for any other analyses.

Geometric morphometrics – Lower pharyngeal jaw shape

Data collection

Lower pharyngeal jaws were excised via the operculum under light microscope. Any remaining soft tissue was removed by hand, and PHJs were dried and mounted on 1-mm scale grid paper for photography. Photographs were taken using a Nikon SM21000 light microscope at x15-30 magnification levels.

Data analysis

Digital images were processed in tpsUtil v 1.58 and landmarks were digitised using tpsDig2 v 2.17. A set of 28 landmarks was initially digitised, comprising 6 true landmarks and 22 semilandmarks describing the outline of the PHJ bone (Figure 5.3). The semilandmarks were subjected to a sliding process in tpsRelw v 1.54 (10 iterations) using the minimum bending energy criterion to minimise differences in landmark placement along the curve. The use of 22 semilandmarks allowed for the

curvature of each section of the LPJ outline to be best captured. However, to avoid over-representation of semi-landmarks vs. true landmarks and to avoid the use of substantially more landmarks than specimens available (complicating statistical inference) the number of data points was subsequently reduced. Thus, after sliding, semi-landmarks were pruned to six paired semilandmarks. The retained semilandmarks were thereafter treated as landmark data, and combined with original landmarks to form a dataset of 12 landmarks. Data was imported to MorphoJ 1.05f and after accounting for object symmetry, analyses were the same as for the body shape data described above. As it is not possible to account for landmark symmetry (paired data) in PAST, one landmark for each pair was removed before conducting NPMANOVA on the pharyngeal jaw data.

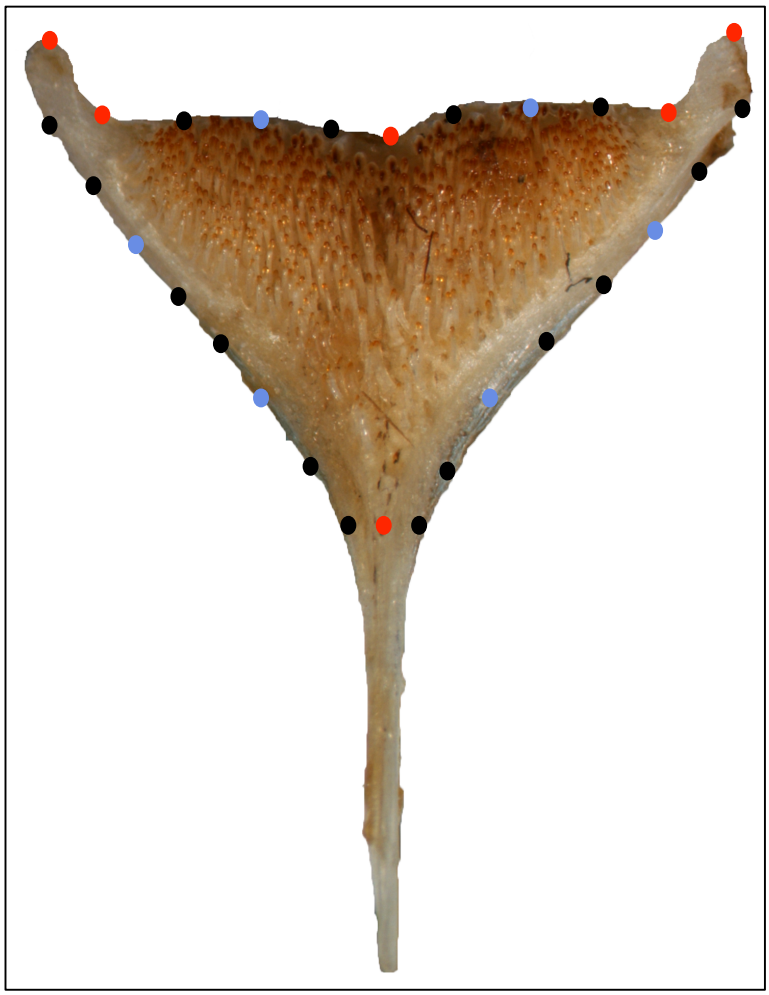


Figure 5.3. Lower pharyngeal jaw landmarks for morphometric analysis.

Red circles: six true landmarks; *blue circles*: the six retained slid-semilandmarks treated as landmark data; *black circles*: semilandmarks discarded after iterative sliding procedure.

Covariation between datasets

The restriction-site associated DNA (RAD) genomic dataset (chapter three) was used to compare genomic and ecomorphological differentiation. Uncorrected pairwise p-distances between individual specimens from the RAD dataset were calculated in the package ape (Analyses of Phylogenetics and Evolution; Paradis *et al.* 2004).

For comparisons requiring a phylogenetic tree, the reduced-taxa phylogeny (generated with RAxML using the full alignment including invariant sites, chapter three) was used for these analyses, and so not all RAD populations are included for comparison (in particular, populations from sites 12 and 24 are excluded). Distances between populations (terminals) in the Maximum Likelihood (ML) tree were calculated using the cophenetic.phylo() function in R v 3.1.2 (R Development Core Team 2015). The ML tree from the RAD data was prepared for plotting in morphospace by pruning the phylogeny to include only one individual per population (species/site) using Mesquite v 2.75 (Maddison & Maddison 2011). The individual retained per population was selected according to the quality of RAD sequences; keeping the individual with the highest number of filtered calls for each population. Shape reconstruction for ancestral nodes in the molecular phylogeny was conducted in MorphoJ using squared-change parsimony weighted by the degree of molecular change on the respective branches of the tree.

In order to visualise differentiation of genomic data compared to that of the morphometric data, RAD data (chapter three) was re-analysed for this chapter to visualise the data in a principal component analysis. As missing data can influence the results of genetic PCA (if data is missing non-randomly across the data set), only RAD sites with data for all individuals (i.e., no missing data), and that were unlinked (imposing minimum distance of 500kb between sites) were included in the PCA, leaving a total of 818 SNPs across all *Alcolapia* samples. The analysis was conducted in R v 3.0.3 using packages adegenet (Jombart 2008) and ade4 (Dray & Dufour 2007).

Covariate analysis was conducted between the morphometric and stable isotope datasets using a partial least squares regression in MorphoJ. Covariation of stable isotope and genetic distance was tested using Mantel tests in the ade4 package in R (Dray & Dufour 2007), comparing pairwise distance between individuals within each site for stable isotope (Cartesian co-ordinates) vs. RAD (uncorrected p-distance) datasets, by species. Geographic comparisons were not possible between populations for the stable isotope data due to the lack of baseline

data. Simple and partial Mantel tests were conducted using the ecodist package in R (Goslee & Urban 2007) to test matrix covariation of body shape morphometric and genomic datasets (pairwise F_{ST} values) while controlling for geography. All tests for morphological covariation were at the population (intraspecific level only), so included different numbers of populations per species. Site 17 and the upturned morph from site 15 were excluded from the *A. alcalica* analysis. Only overall body shape for GMM data was considered for these analyses, as PHJ data was available from only a subset of populations for each species. Mantel tests were all conducted with 10,000 permutations.

As comparison of morphological distance to genetic distance suggested possible cases of body shape convergence within the dataset (see Results), this relationship was explored statistically using the R package SURFACE (SURFACE Uses Regime Fitting with Akaike Information Criterion (AIC) to model Convergent Evolution) (Ingram & Mahler 2013). SURFACE uses phylogenetic and phenotypic information to model lineages undergoing shifts to adaptive peaks on a macroevolutionary landscape, but does not use *a priori* information regarding which lineages correspond to particular peaks (Ingram & Mahler 2013; Mahler *et al.* 2013). SURFACE consists of a forward phase starting using the Ornstein-Uhlenbeck model to add regimes to a Hansen model with all lineages at a single peak in trait space, and fits increasingly complex (multipeak) models based on Akaike information criterion (AICc), placing the adaptive peak shift on whichever branch of the phylogeny most improves the AICc at each step. The backward step collapses the number of regime shifts to assess if reduction in parameter number outweighs any decrease in log-likelihood (Ingram & Mahler 2013). The models have been shown to have good power to detect convergence, as well suitable error rates when there are adaptive peak shifts but no convergence (Mahler *et al.* 2013), although these tests were conducted on moderately sized radiations that are substantially larger than the radiation considered here. Although most recent implementations of the SURFACE method have used time-calibrated Bayesian maximum clade credibility trees to account for phylogenetic uncertainty (e.g., Mahler *et al.* 2013; Davis *et al.* 2014; Bravo *et al.* 2014; Arbour & Lopez-Fernández 2014; Astudillo-Clavijo *et al.* 2015), previous studies have also used maximum likelihood calibrated phylogenies (Grundler & Rabosky 2014), and the original methods description included testing on a UPGMA phylogeny with root age scaled to a single time point (Ingram & Mahler 2013).

As SURFACE requires an ultrametric tree as input, a chronogram was fitted to the ML phylogeny from chapter three using the *chronos* function of the ape package

in R. The chronogram was fitted using the discrete model of substitution rate variation among branches, a lambda value of 1, the default control settings, and with root node age scaled to 1. As a time-calibrated phylogeny was not available, the SURFACE results (i.e., patterns of regime shifts) cannot be assigned a time scale. Where populations were excluded from analyses, these nodes were pruned in Mesquite 3.02 (Maddison & Maddison 2014) before fitting the chronogram in R. Size-corrected PC scores for population means were exported from MorphoJ and PC1 and PC2 used as input for SURFACE analysis. Where populations were excluded from analysis, PCA was conducted separately on the reduced dataset to generate PC scores for the focal populations alone.

Results

Trophic niche differentiation

Stable isotope analysis

A total of 360 individuals were analysed for stable isotope change across all populations (mean: n=16 per population). Biplots of individual $\delta^{13}C$ and $\delta^{15}N$ isotopic values for all Lake Natron sites at which species occurred sympatrically revealed that *A. alcalica* exploit significantly different isotopic niches than both *A. latilabris* and *A. ndalalani*, which overlapped in niche space at all sites (Figure 5.4A). *Alcolapia alcalica* was consistently $\delta^{13}C$–enriched relative to the other two species, indicating that *A. alcalica* is feeding on a different food source, but at the same trophic level, as the other two species. The within-site ANOVA tests between species demonstrated that significant differences were found within the $\delta^{13}C$ values for all *A. alcalica* comparisons except at site 17 (Figure 5.4B).

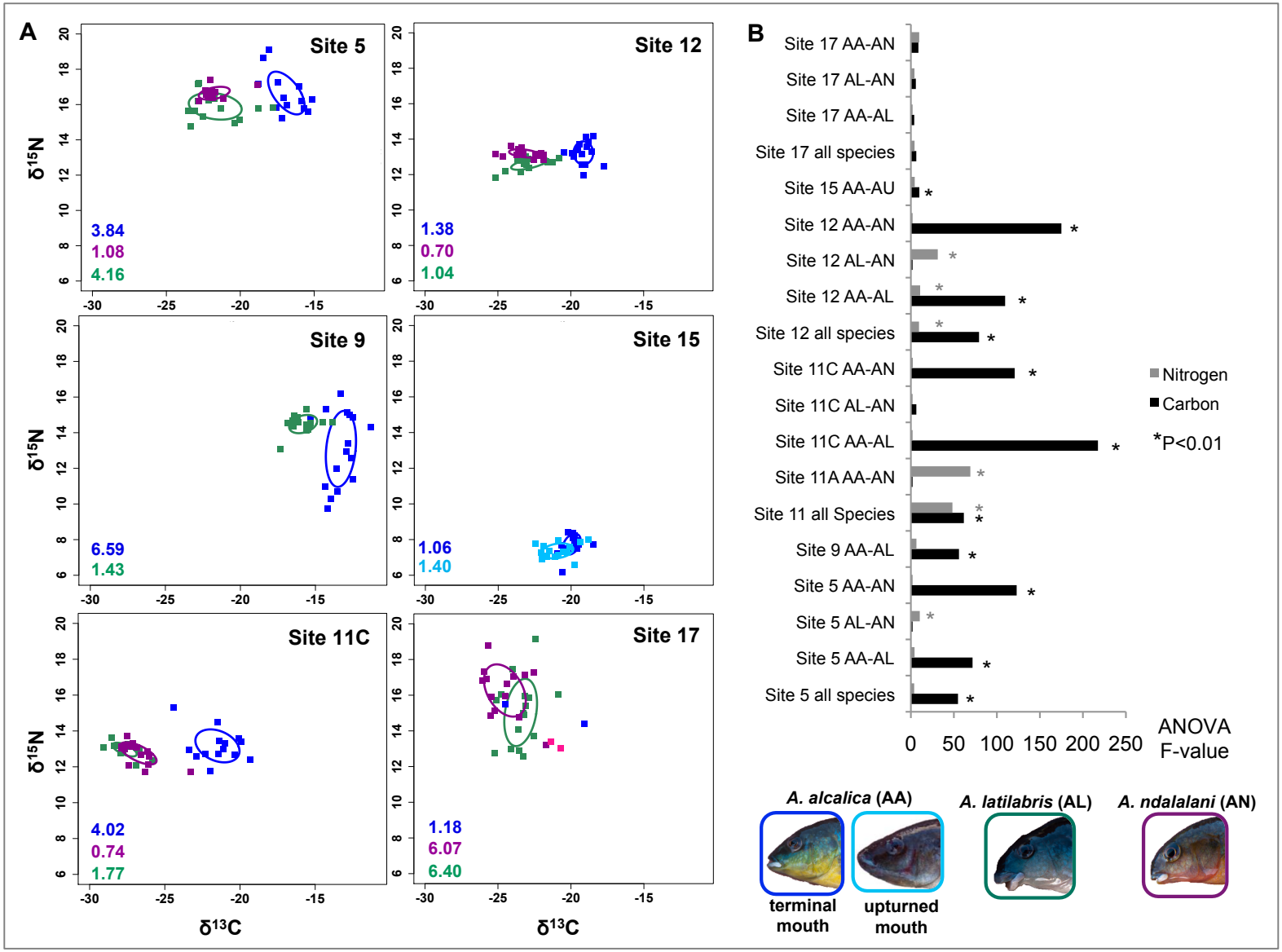

A
Site 5
3.84
1.08
4.16
Site 12
1.38
0.70
1.04
Site 9
6.59
1.43
Site 15
1.06
1.40
Site 11C
4.02
0.74
1.77
Site 17
1.18
6.07
6.40
δ15N
δ13C
B
Site 17 AA-AN
Site 17 AL-AN
Site 17 AA-AL
Site 17 all species
Site 15 AA-AU
Site 12 AA-AN
Site 12 AL-AN
Site 12 AA-AL
Site 12 all species
Site 11C AA-AN
Site 11C AL-AN
Site 11C AA-AL
Site 11A AA-AN
Site 11 all Species
Site 9 AA-AL
Site 5 AA-AN
Site 5 AL-AN
Site 5 AA-AL
Site 5 all species
Nitrogen
Carbon
*P<0.01
0 50 100 150 200 250
ANOVA F-value
A. alcalica (AA)
terminal mouth
upturned mouth
A. latilabris (AL)
A. ndalalani (AN)

Figure 5.4.

Legend for Figure 5.4 (previous page). Stable isotope differentiation of *Alcolapia* populations. A: Biplots of stable isotope values for nitrogen and carbon isotope ratios in delta per mille, for each of the Lake Natron sites where species/morphs occur sympatrically. Values in the left hand corner of each plot represent the ellipse area for each species. Data for site 17 has been separated for *A. alcalica* (blue) and *A*. aff. *ndalalani* (light pink) individuals, as identified by genomic analysis in chapter three. As each of these groups contained only two individuals, it is not possible to plot an ellipse. **B:** Results of the within-site ANOVA tests for stable isotope values. Significant differences are observed within the carbon isotope values for all *A. alcalica* comparisons except at site 017.

Although the ANOVA tests found significant differences between certain population means of *A. latilabris* and *A. ndalalani* for either $\delta^{13}C$ or $\delta^{15}N$ (never both at the same site), it seems unlikely that these differences are biologically relevant given the very small differences in means and remaining overlap of individual variation between species. Certainly, the differences in mean $\delta^{15}N$ between these two species (site 005: 0.740‰; site 12: 0.579‰) is less than that typically defining a trophic level (2-4‰; Perkins *et al.* 2014). While differences in $\delta^{15}N$ may reliably indicate trophic level, $\delta^{13}C$ may not reliably distinguish different feeding sources, as where different basal sources exhibit the same $\delta^{13}C$ signal, these differences would be maintained throughout the food chain (Perkins *et al.* 2014).

Stomach contents and gut length

A total of 121 individuals were measured for gut length and 122 analysed for stomach contents (94% individuals were included for both analyses, with 6% included in only one analysis). Individuals were included from three sympatric populations for the Lake Natron species (sites 5, 11 and 12), and from the Lake Magadi populations (sites 18 and 21) for *A. grahami*. Intraspecific variability between populations is discussed in chapter six. Mean relative gut length ($\log_{10}$ gut/body length) was significantly different between *A. alcalica* and *A. grahami*, as well as between *A. alcalica* and *A. ndalalani* (Figure 5.5). However, the three Lake Natron species all overlapped considerably, so the mean difference is likely only biologically relevant in the *A. alcalica* to *A. grahami* comparison. The three Lake Natron species all exhibited a greater range of values than *A. grahami*, but this is probably due to the *A. grahami* dataset including data from fewer populations and samples.

Stomach contents analysis suggested that all species were mainly herbivorous, with limited contribution from other sources (Figure 5.5).

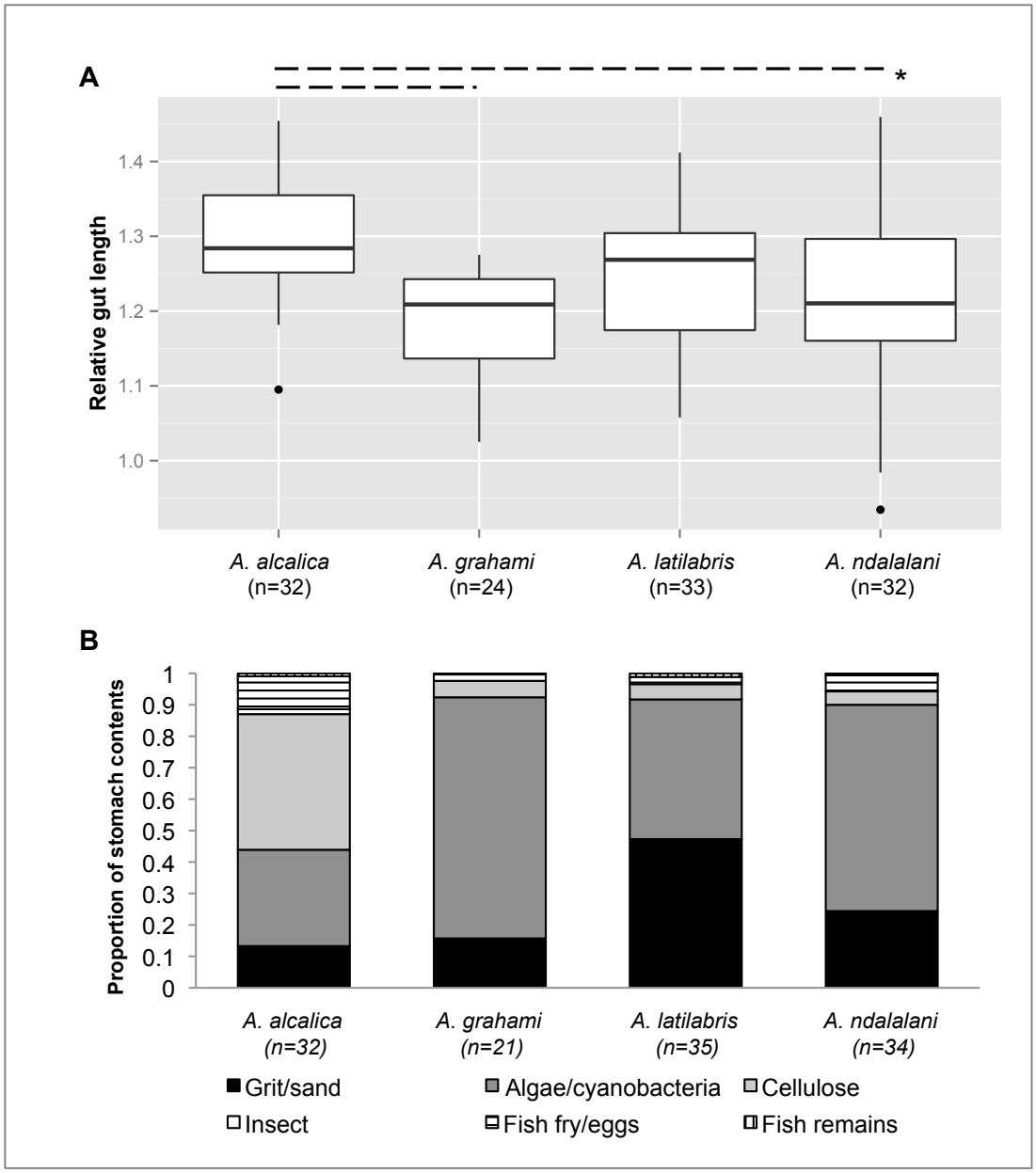


Figure 5.5. Gut length ratio and stomach contents analysis by species.

Data are included for three Lake Natron populations (sites 05, 11 and 12) and two Lake Magadi populations (sites 18 and 21). A) Gut length to body (SL) ratio. Asterisk indicates pairwise ANOVA comparisons significant at α <0.05 following sequential Bonferroni correction. B) Stomach contents by proportion. Total specimen numbers differ between analyses as individuals where gut disintegrated during uncoiling or those with entirely empty stomachs were excluded.

In stomach contents analysis, a substantial proportion (43%) of *A. alcalica* diet was accounted for by plant material (cellulose), with a smaller proportion (30%) of algae and cyanobacteria, while all other species exhibited a major proportion of diet based on algae (44-77%) with only minor contributions of plant material (4-5%). Comparing overlap between species diet using Schoener's index indicated that the diet of

A. alcalica was different from all other species, but comparisons among all other species indicated substantial overlap (using a threshold of 0.6 for overlap) (Figure 5.6).

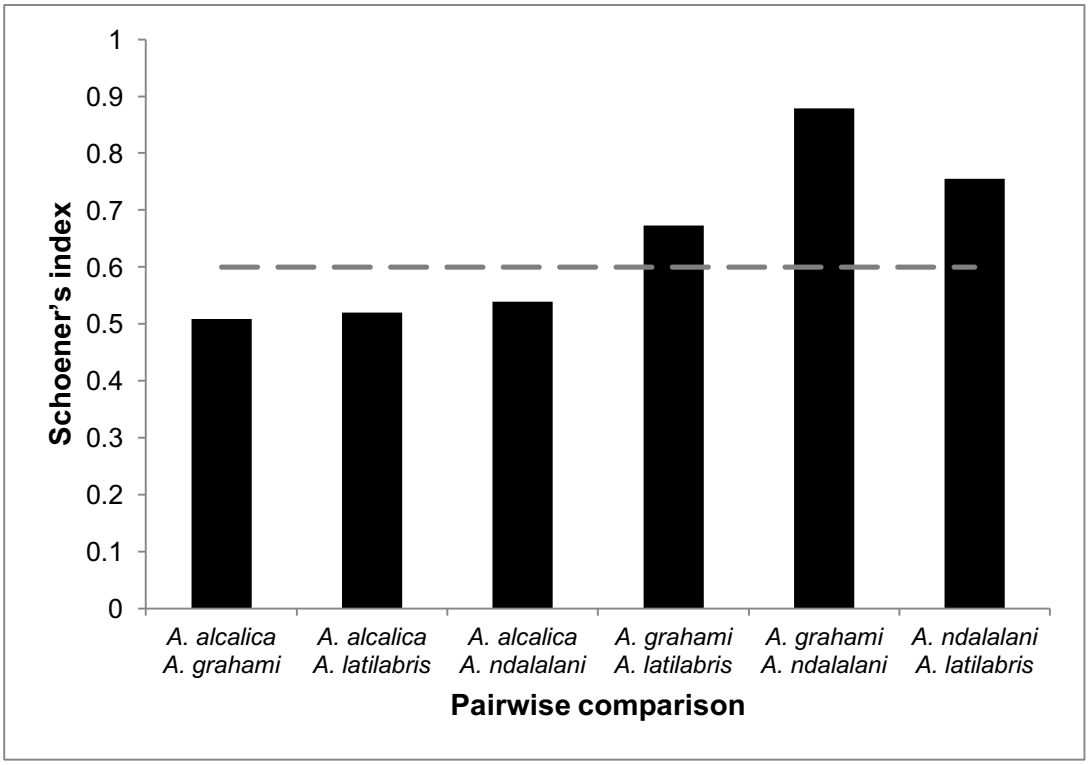


Figure 5.6. Pairwise comparison for Schoener's Index of dietary overlap.
Dashed grey line indicates threshold above which comparisons indicate substantial dietary overlap.

As well as differences in relative proportions of cellulose and algae components, *A. latilabris* exhibited considerably higher proportion of grit than other species. Furthermore particle size of the sand/grit component differed between Lake Natron species with *A. latilabris* having a substantially larger particle size than *A. ndalalani* (pers. obs.), which may suggest differences in foraging mode.

Geometric Morphometrics

Variation of body shape

A total of 567 individuals were included for the geometric morphometric analysis of body shape in the analysis including only populations for which genomic RAD data was available. Of the populations that were RAD sequenced, site 17 *A. alcalica* samples were excluded from morphometric analysis, as only 4 samples were available, two of which had been identified as possible hybrids (chapter three) and

so were deemed unsuitable for species-level morphological analysis. Of the remaining two samples deemed to be valid *A. alcalica* from site 17, only one voucher sample was available for morphology, so to avoid a group size of n=1, the population was removed from the analysis. Although not included for morphometric analysis, photographs of the site 17 *A. alcalica* and *A. aff. ndalalani* are included in Appendix Figure 5A.1 for information.

Following Procrustes superimposition, the spread of coordinates within the dataset varied across landmarks (Figure 5.7), with cranial landmarks generally showing most spread around the consensus landmark configuration, and the postcranial landmarks showing a smaller degree of variation. Of the cranial landmarks, the two orbit points (centre and dorsal margin, landmarks 15 and 16; Figure 5.2) exhibited comparatively low levels of variation, while the two anterior lip landmarks (anterior snout tip and anterior tip of maxilla, landmarks 12 and 13; Figure 5.2) showed the most variation of all landmarks across the dataset.

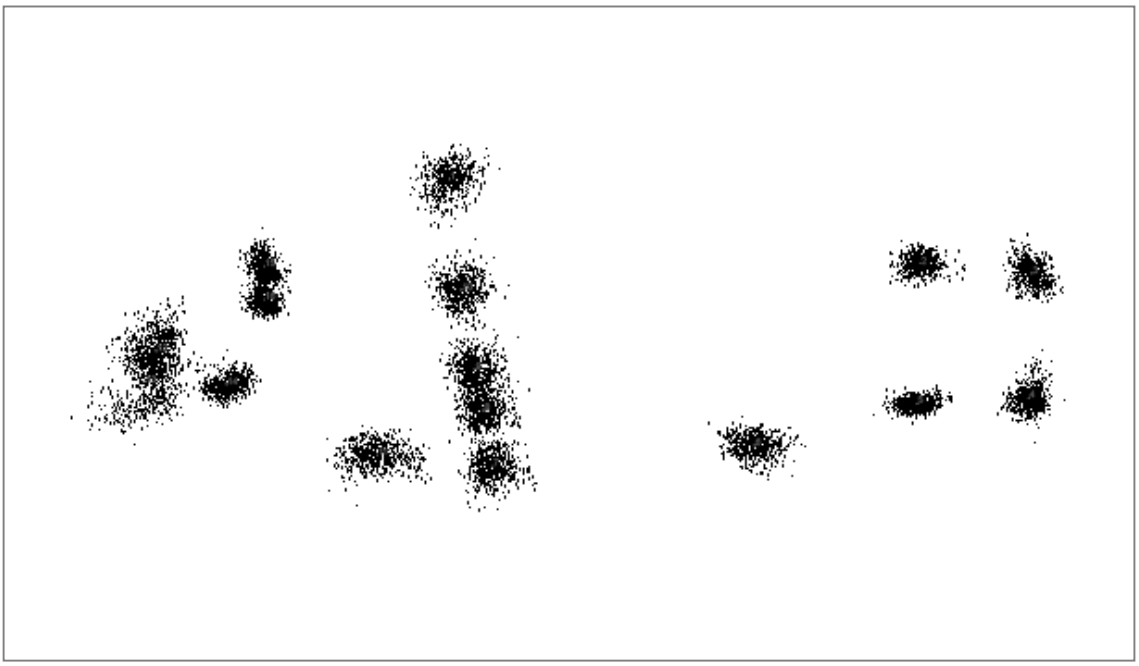


Figure 5.7. Landmark variation for geometric morphometrics of body shape. Landmark scatter for Procrustes-aligned co-ordinates of all samples included in the geometric morphometric analysis (n=567).

Furthermore, while most landmarks appear to show either an equidistant scatter in all directions around the consensus landmark coordinates or a unidirectional scatter along the anterior-posterior axis, landmark 13 appears to show two distinct patterns of scatter both antero-dorsally and antero-ventrally from the consensus (Figure 5.8), suggesting two distinct angles of shape variation.

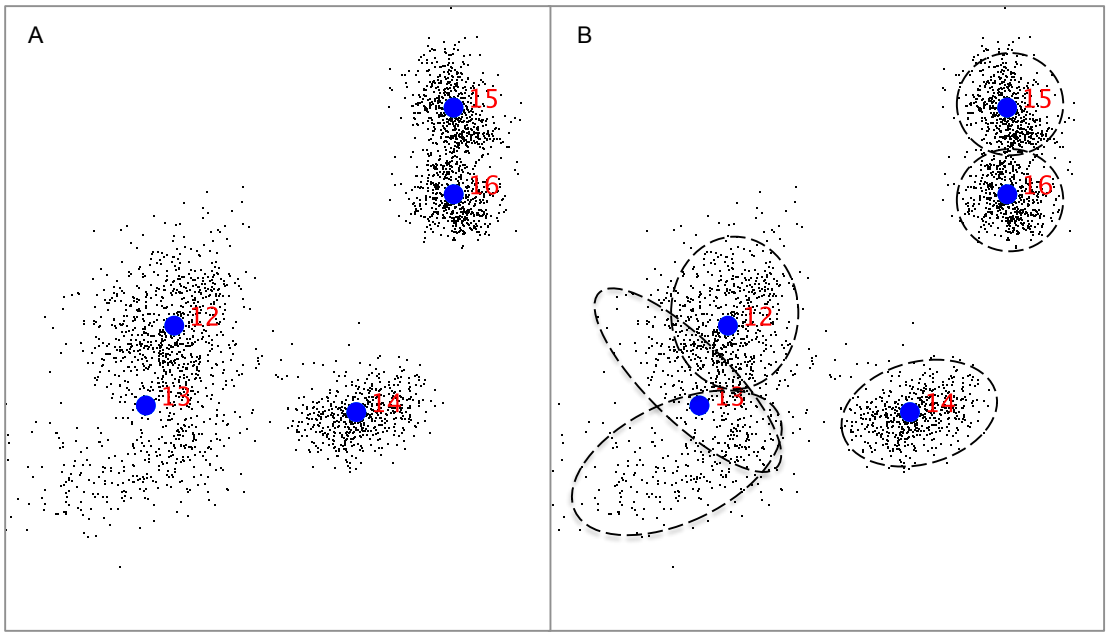


Figure 5.8. Variation in cranial landmark scatter. A) Close-up of landmark coordinate scatter around consensus mean shape for cranial landmarks 12-16. Blue circles represent consensus configuration across the entire dataset. B) Ellipses describing approximate scatter variation (placed by eye) in direction related to consensus, where landmark 13 appears to exhibit two patterns of scatter, with very few landmarks exhibiting intermediate coordinates.

For comparison, plotting the Procrustes-aligned X and Y coordinates for landmark 13 and 16, landmark 13 Y coordinates exhibit a bimodal distribution, while landmark 13 X coordinates, and both X and Y coordinates for landmark 16, exhibit a unimodal distribution (Figure 5.9). As landmark 13 represents the anterior snout tip, this pattern of distribution indicates that the sample set largely includes individuals with either an upturned or downturned lip juncture, but very few exhibiting an intermediate profile, while the variation in the horizontal plane (X coordinates) tends towards a normal distribution across all samples in the dataset for this landmark.

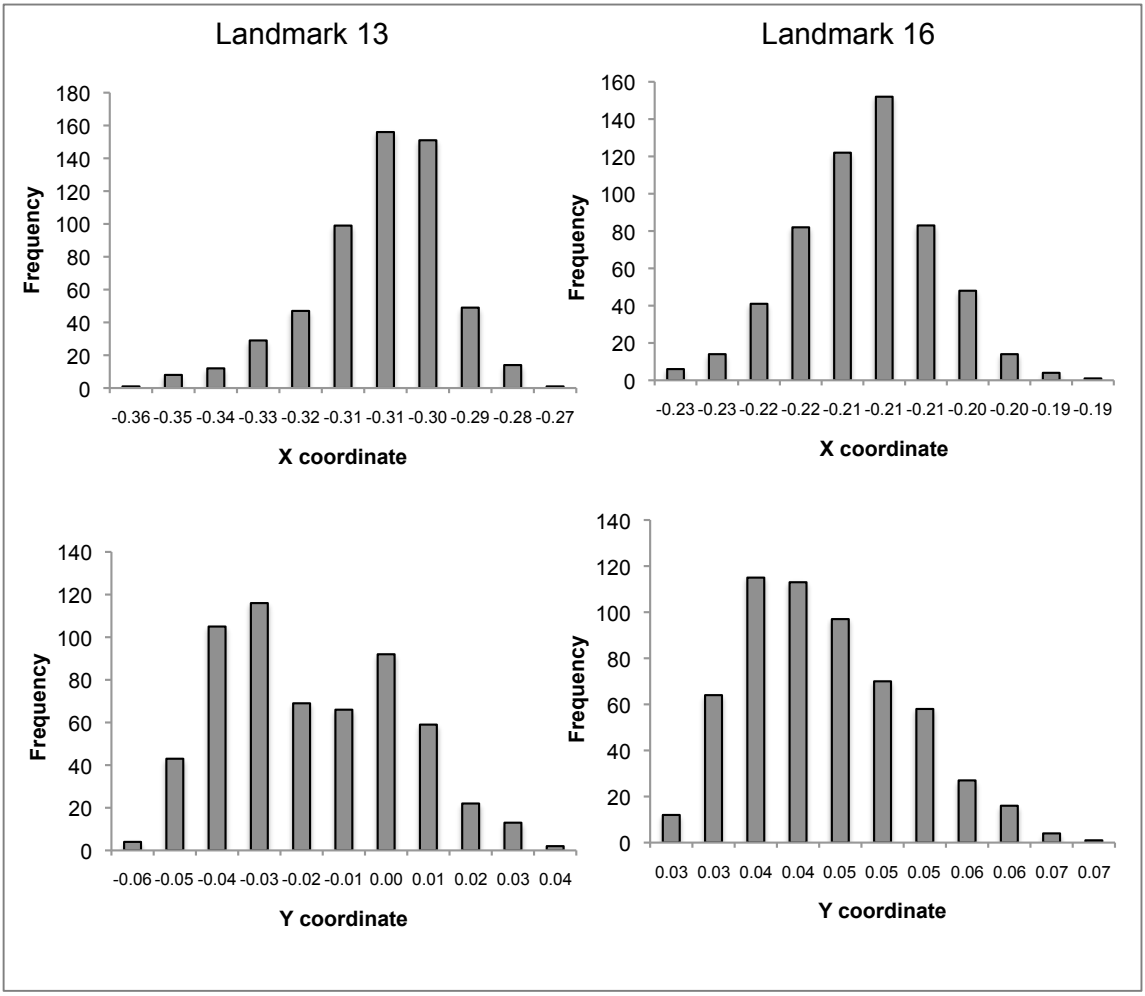


Figure 5.9. Histogram of Procrustes-aligned X and Y co-ordinates for two cranial landmark points. While landmark 16 exhibits a unimodal distribution in both dimensions, landmark 13 exhibits a bimodal distribution of Y coordinates.

Landmark 11 (anteroventral tip of pectoral girdle, Figure 5.2) also exhibited a large degree of variation, most noticeably in the horizontal plane (Figure 5.7). This may most prominently reflect differences in head length relative to body length, variation that has previously been recorded between these species, being notably longer in *A. latilabris* than in other Lake Natron species (Seegers & Tichy 1999). This bimodality of positional scatter for landmark 13 was not seen when Procrustes alignments were performed for single-species datasets used for intraspecific analysis (Appendix Figure 5A.2).

Regression of Procrustes coordinates against centroid size was significant (10,000 permutations; $P<0.0001$) and size accounted for 5.60% of the variation within the dataset, thereafter size-corrected results (regression residuals) were used for all downstream analyses. In the PCA on the Procrustes-fitted and sized-corrected residuals across the entire dataset, the first three PCs accounted for 67% of the variation within the dataset (all other PC Eigenvalues accounted for 5% or less variation each; total of 28 PCs; Appendix Table 5A.1). When PCA was conducted on group means of each population (species/site) increased variance was explained by each eigenvalue.

PC1 and PC2 demonstrated the greatest variation, with most difference indicated by head and lip shape and minor difference in body depth. The data clustered by species with minimal overlap, with *A. alcalica* closest to *A. grahami* in morphometric space (Figure 5.10).

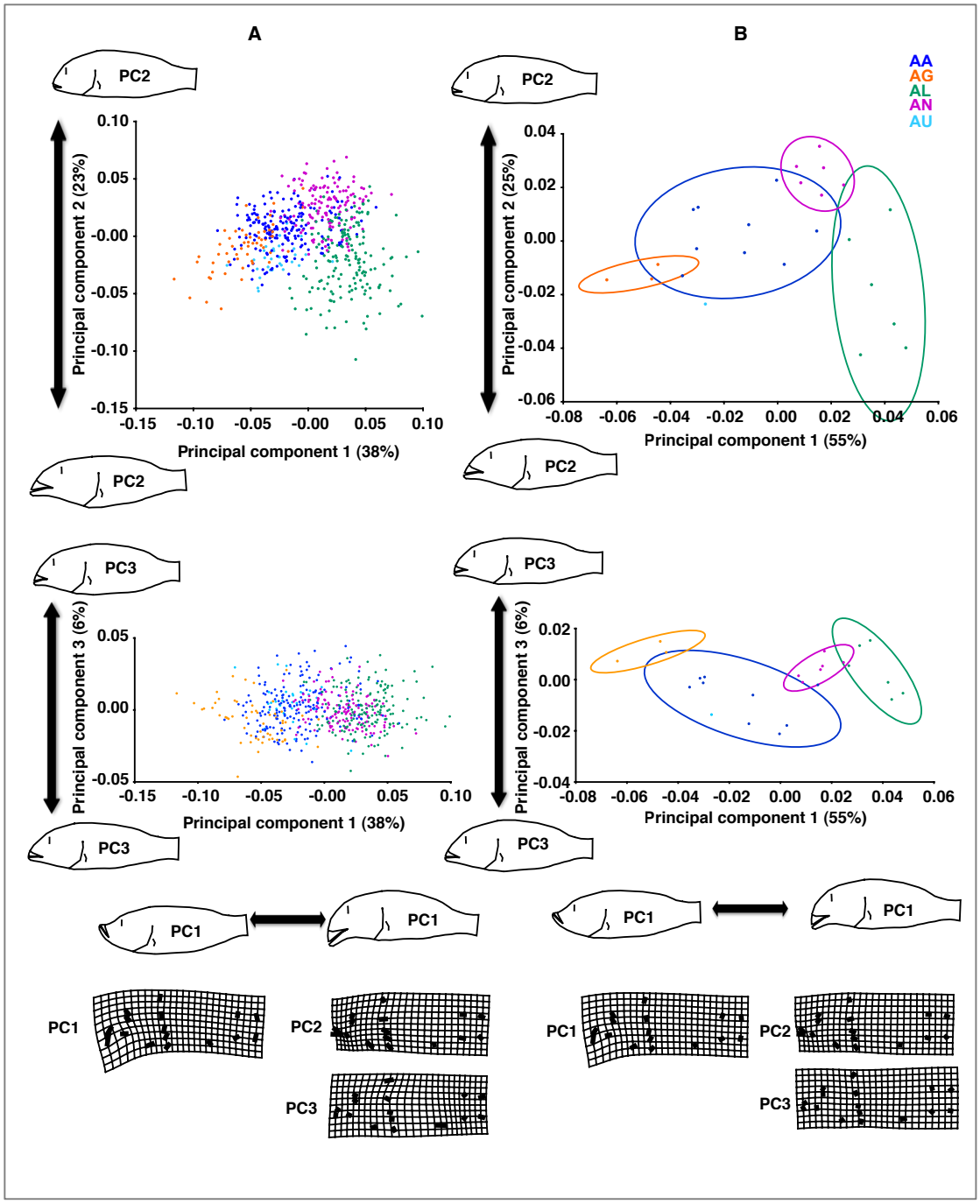


Figure 5.10. Principal components analysis (PCA) of body shape variation across all sampling sites (n=567 individuals) of PC1 vs. PC2 and PC1 vs. PC3 for A) all individuals; B) population means for species grouped by sampling site. Outline shape drawings represent the shape at the minimum and maximum extent of data along each PC axis. Warped transformation grids show maximum change from consensus shape along the positive axis only. Ellipses represent the variation of each group, drawn as equal frequency ellipses at a probability of 0.9 (i.e., such that 90% of all variation of the sample is found within the ellipse area).

As may be expected, the shape changes for all data analysed individually vs. pooled by species and sampling site were very similar, however those for the mean values showed less extreme variation. Repeating the PCA with *A. alcalica* subdivided into two groups forming the 'northern' (sites: 6, 15, 19) and 'southern' clades identified in phylogenetic analysis (chapter three) exhibited very tight clustering in morphospace of the northern populations, which overlapped with *A. grahami*, but much wider variation of morphology between the southern populations, which were distinct from *A. grahami* but overlapped with *A. ndalalani* (Figure 5.11). As such, the southern populations appear to show closer morphological affinity to *A. ndalalani,* with which they occur in sympatry.

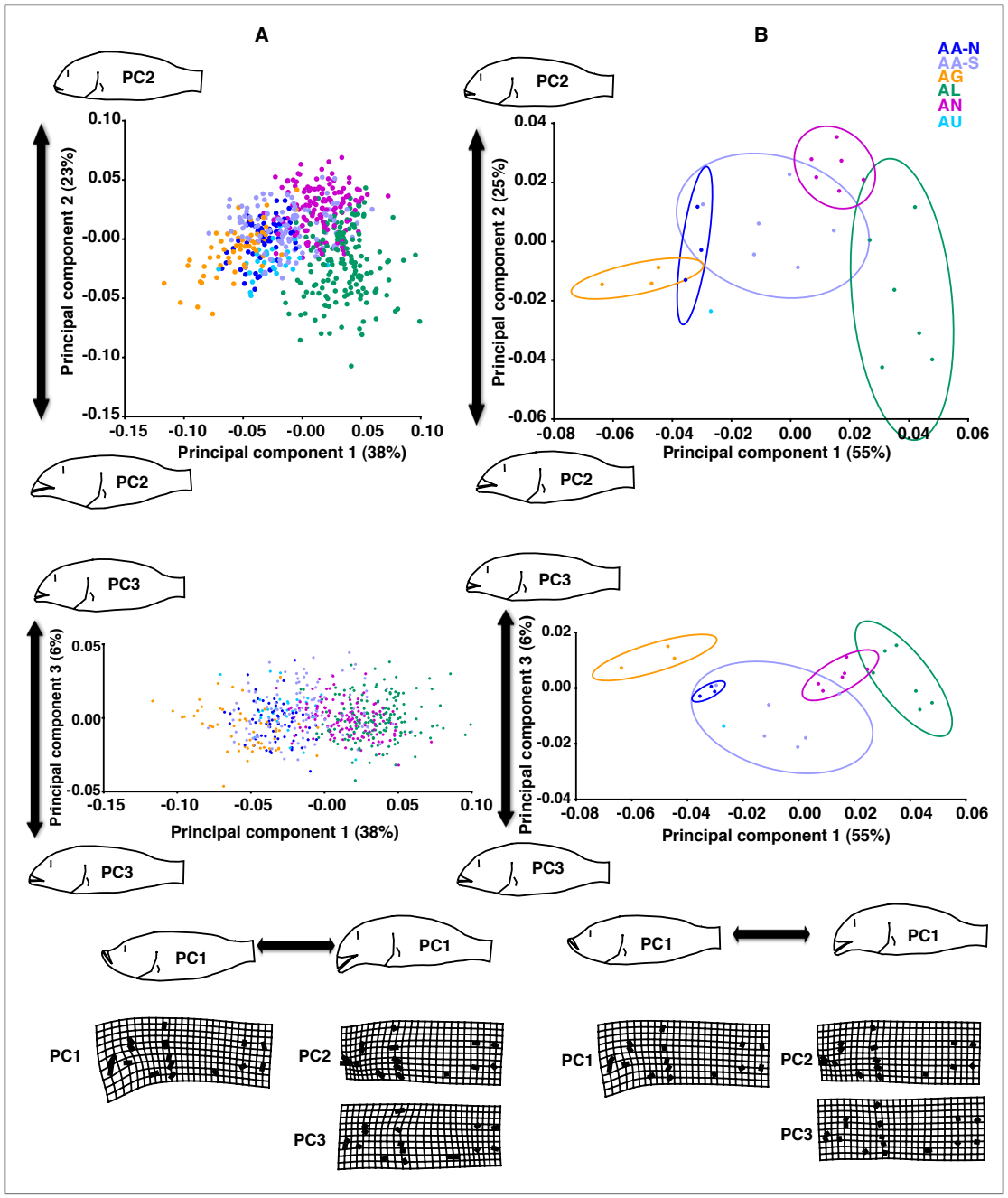


Figure 5.11. Principal components analysis of *Alcolapia* populations.

Analysis as for Figure 5.10, but with *A. alcalica* data points coloured by clade for northern and southern populations.

While *A. grahami* and *A. alcalica* overlap in PC1 (mouth orientation) and PC2 (snout/head length), they are differentiated by PC3 that describes body depth and length (Figure 5.10; Figure 5.11).

All species pairwise comparisons were significantly different in CVA analysis (all P<0.0001 from 10000-round permutation test on Mahalanobis distance; Tables 5.2 and 5.3), although not for the *A. alcalica* upturned mouth morph when pooled by site. However, this is likely due to low statistical power as the upturned mouth morph

was only found from site 15 of the sites surveyed (and to my knowledge has not previously been described from the literature).

Table 5.2. Interspecies-distances from CVA of body shape.

Species pooled by site, for *A. alcalica* analysed as a single group and divided into northern and southern clades**.**

	AA	AA-N	AA-S	AG	AL	AN	AU
AA (n=9)	**-**	**-**	**-**	**0.05***	**0.06***	**0.04***	**0.03**
AA-N (n=3)	**-**	**-**	**0.03***	**0.03**	**0.07***	**0.06***	**0.03**
AA-S (n=6)	**-**	**8.33***	**-**	**0.05***	**0.05***	**0.04***	**0.04**
AG (n=3)	**12.40***	**10.58**	**13.36***	**-**	**0.09***	**0.08***	**0.04**
AL (n=6)	**23.38***	**23.08***	**22.67***	**33.49***	**-**	**0.05***	**0.07**
AN (n=6)	**14.28***	**12.77***	**15.09***	**21.71***	**18.54***	**-**	**0.07**
AU (n=1)	**12.94**	**12.58**	**13.78***	**9.16**	**34.48***	**24.36***	**-**

*Pairwise comparisons significant at P<0.05 (10,000 permutation rounds).

Species comparisons values based on a CVA containing only five groups, while all AA-N and AA-S comparisons based on a separate CVA containing six groups with AA subdivided into two groups.

Below diagonal: Mahalanobis distances among groups; Above diagonal: Procrustes distances.

Table 5.3. Interspecies-distances from CVA of body shape, all individuals.

	AA	AA-N	AA-S	AG	AL	AN	AU
AA (n=181)	**-**	**-**	**-**	**0.04***	**0.06***	**0.04***	**0.03***
AA-N (n=65)	**-**	**-**	**0.02***	**0.03***	**0.07***	**0.06***	**0.03**
AA-S (n=116)	**-**	**2.25***	**-**	**0.05***	**0.06***	**0.04***	**0.04***
AG (n=56)	**3.16***	**2.85***	**3.71***	**-**	**0.10***	**0.08***	**0.04***
AL (n=166)	**6.18***	**7.08***	**6.07***	**7.57***	**-**	**0.05***	**0.07***
AN (n=134)	**3.44***	**4.37***	**3.27***	**4.95***	**5.39***	**-**	**0.07***
AU (n=30)	**3.60***	**3.25***	**4.20***	**4.78***	**8.05***	**5.91***	**-**

*Pairwise comparisons significant at P<0.0001 (10,000 permutation rounds).

As for Table 5.2, analysis conducted in two separate CVAs – all values reported from the 5-group CVA, except for comparisons including AA-N and AA-S from the 6-group CVA.

Below diagonal: Mahalanobis distances among groups; Above diagonal: Procrustes distances.

The pooled site CVA results for Mahalanobis distance are also presented in graphical format in Figure 5.12. Across all comparisons, *A. latilabris* consistently exhibited the most differentiation, with the greatest inter-species morphological distance. The greatest distances were observed between *A. latilabris* vs. *A. grahami*, and *A. latilabris* vs. *A. alcalica* upturned-mouth morph. As indicated by the PCA analysis, the northern and southern clades of *A. alcalica* were significantly

morphologically differentiated. While the *A. alcalica* upturned-morph was not differentiated when *A. alcalica* was treated as a single group, when analysed considering the northern and southern *A. alcalica* clades, it was significantly differentiated from the latter clade. The smallest pairwise distance across all comparisons was between the upturned mouth *A. alcalica* morph and *A. grahami* (Table 5.2; Figure 5.12).

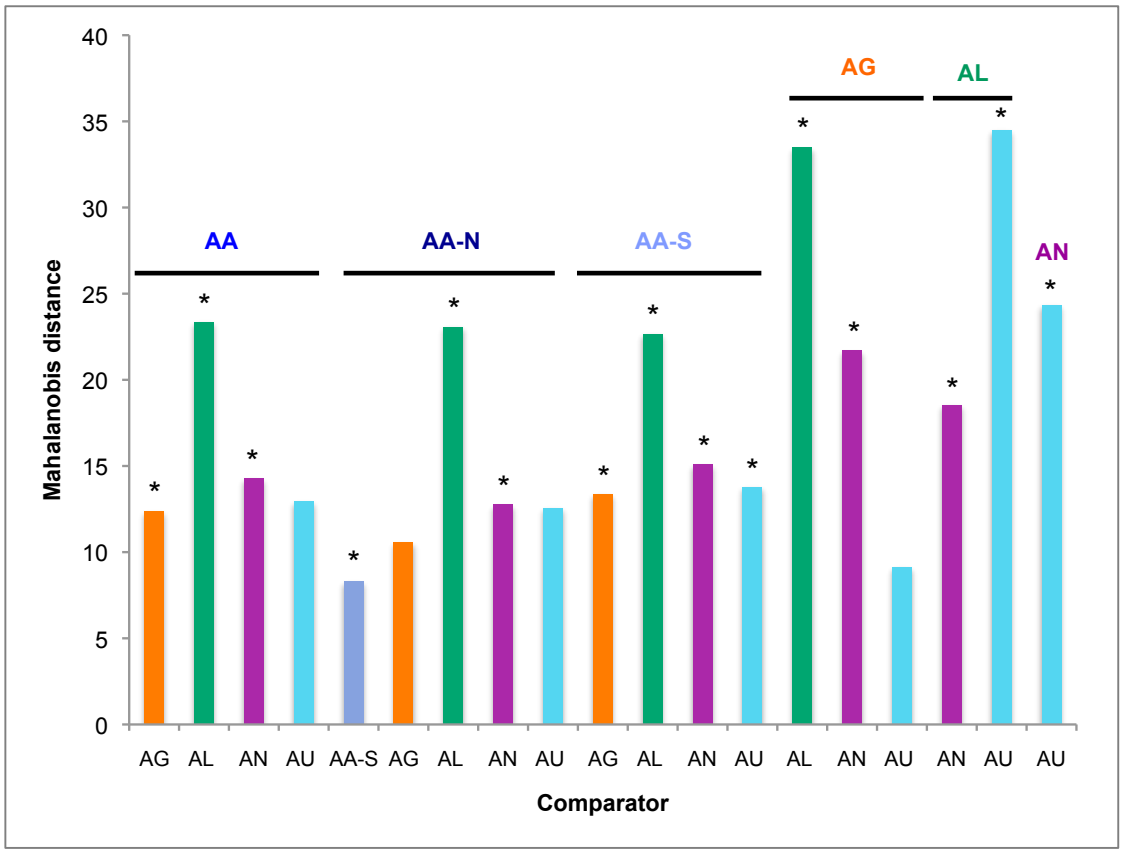


Figure 5.12. Pairwise comparisons of Mahalanobis distance of body shape.
Distances generated from canonical variate analysis. For each pairwise comparison, the first comparator is indicated above the bars and the second on the x-axis. *significant at P<0.05; 10,000 permutations. *Alcolapia latilabris* (green) displays the highest distance values in all comparisons, with the greatest distances between *A. latilabris* vs. *A. grahami*, and *A. latilabris* vs. *A. alcalica* upturned-mouth morph.

Canonical variate analysis maximises the separation of predefined groups (maximising differentiation between group means relative to variation across the group). As such, when plotting the canonical variates, as might be expected, the groups show similar patters of variation to the analysis of the PCA (Figures 5.10 and 5.11), but display more separation of the groups (Figure 5.13). *Alcolapia latilabris*, which exhibits the greatest within-group variation, is the most clearly separated of

the groups, while *A. alcalica* clades/morphs and *A. grahami* still exhibit a substantial degree of overlap. *Alcolapia ndalalani* exhibits a somewhat intermediate dispersion, being generally separated from the other groups, but positioned at the intersections of the axes on which *A. latilabris* and *A. alcalica*/*A. grahami* groupings are situated. As might be expected, given that the analysis provides ordination to minimise variation within groups while maximising that between them, all groups display tighter grouping and lower variation as demonstrated by the probability ellipses. Colouring the data points considering the northern and southern *A. alcalica* clades (Figure 5.13; right hand panel) demonstrates that the substantial remaining overlap between *A. alcalica* and *A. grahami* is mostly due to overlap from the northern *A. alcalica* clade, while the southern *A. alcalica* clade still exhibits a small degree of overlap with *A. ndalalani*.

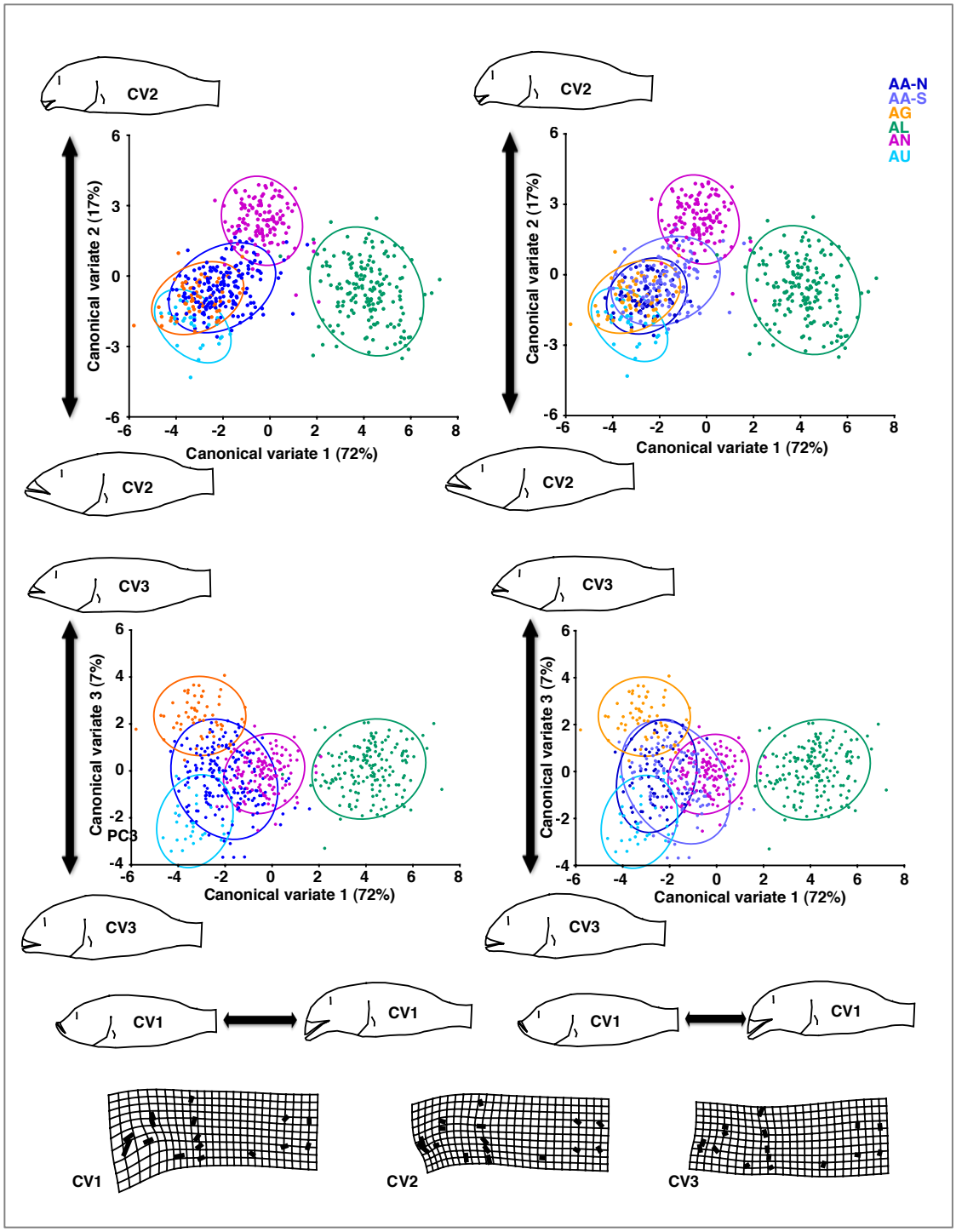


Figure 5.13. Canonical variate analysis by species/morph.

Alcolapia alcalica is coloured as one species (right-hand panel) or as two separate north and south clades (left-hand panel). While *A. grahami* and *A. alcalica* overlap almost entirely in the first two axes describing mouth orientation and snout length, they are differentiated in CV3, describing differences in body depth.

Discriminant function analysis (DFA) conducted on individual data grouped by species and clade significantly separated all groups (P<0.0001). Comparing the

shape changes between groups indicates the large contribution of head and oral trophic morphology to overall shape variations within *Alcolapia*, while body depth also noticeably differs between certain populations (Figures 5.14 and 5.15). The main differences in shape variation between the northern and southern clades of *A. alcalica* include mouth orientation and snout length, with the northern clade exhibiting a more upturned mouth position and longer head and snout length (Figure 5.14). In the respective comparisons of the *A. alcalica* clades with other *Alcolapia* species, the northern clade appears more similar morphologically to *A. grahami* while the southern clade exhibits much more pronounced differences in mouth orientation, snout length and body depth. Whereas, the converse is true in comparisons with *A. ndalalani*, where the southern *A. alcalica* clade is much closer in shape of mouth and body depth, and rounded forehead, while the northern clade has a more upturned mouth, longer head and more gently sloping forehead.

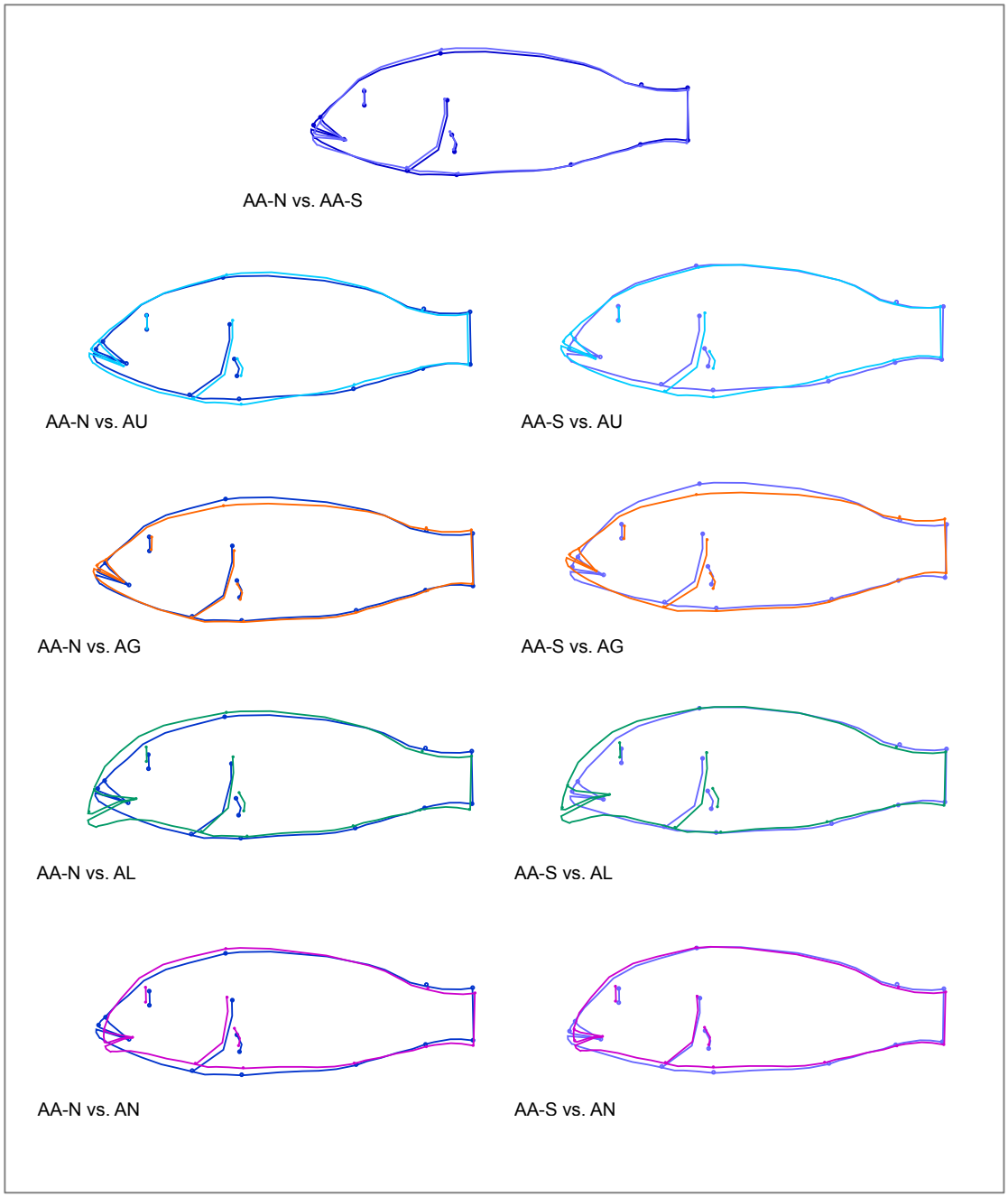


Figure 5.14. Morphological body shape differences for *A. alcalica* clades.
Pairwise comparisons of outline shape difference shown between *A. alcalica* clades (north/south) and other *Alcolapia* species, generated by discriminant function analysis. The *A. alcalica* northern clade (AA-N; dark blue) exhibits observable differences from *A. alcalica* southern clade (AA-S; mauve), including mouth orientation, snout length and body depth. AG: *A. grahami*; AL: *A. latilabris*; AN: *A. ndalalani*; AU: *A. alcalica* upturned-mouth morph.

In all species comparisons, *A. grahami* exhibits a narrower body, less steeply sloping forehead, and comparatively upturned mouth (Figures 5.14. 5.15). *Alcolapia grahami* and the *A. alcalica* upturned-mouth morph exhibit very similar shape profiles with the same mouth orientation and head length, the main difference being

body depth with *A. grahami* having a narrower body. *Alcolapia latilabris* is the most differentiated based on large differences in mouth morphology in all comparisons, also having a particularly long head and very rounded forehead. Finally, *A. ndalalani* exhibits the bluntest snout and shortest head in all pairwise comparisons.

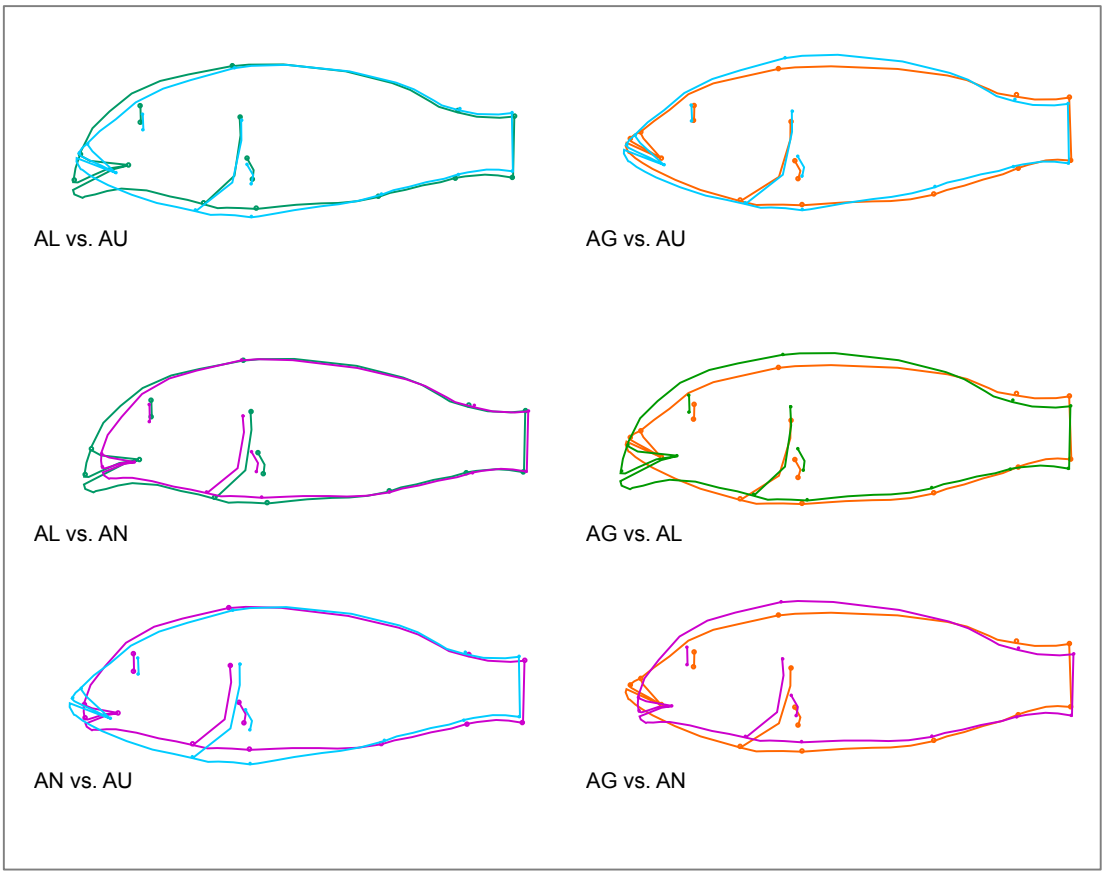


Figure 5.15. Morphological body shape differences for *Alcolapia*.
As for Figure 5.14, pairwise shape differences between species and clades of *Alcolapia*.

Nonparametric methods were also used to validate the statistical significance of the CVA results. All pairwise species comparisons using DFA were significant at P<0.0001, and NPMANOVA results showed a similar pattern to those generated using CVA (Table 5.4).

Table 5.4. Pairwise F-values for NPMANOVA between species and clades.

	AA	AA-N	AA-S	AG	AL	AN	AU
AA (n=181)	**-**	**-**	**-**	**32.94***	**137.80***	**67.75***	**13.25***
AA-N (n=65)	**-**	**-**	**11.92***	**12.70***	**101.70***	**76.92***	**9.91***
AA-S (n=116)	**-**	**3.18***	**-**	**39.67***	**91.68***	**38.56***	**16.28***
AG (n=56)	**4.83***	**3.11***	**3.92***	**-**	**136.80***	**124.00***	**17.47***
AL (n=166)	**9.92***	**6.71***	**7.96***	**6.55***	**-**	**87.45***	**47.66***
AN (n=134)	**7.87***	**5.73***	**6.21***	**5.58***	**8.21***	**-**	**61.94***
AU (n=30)	**4.34***	**2.45***	**4.02***	**2.62***	**5.68***	**54.84***	**-**

*All comparisons significant at P < 0.01 for Bonferroni-corrected P-values (10,000 permutations). Below diagonal: Mahalanobis distances among groups; Above diagonal: Euclidean distances.

When CVA was conducted among all the population groups (i.e., separated by site and species), substantial differentiation was identified within species between sampling sites (Figure 5.16). Although there is some overlap between *A. alcalica* and *A. ndalalani*, and *A. alcalica* and *A. grahami*, *A. latilabris* is almost entirely separated from other *Alcolapia* species. All *A. latilabris* populations are distinct from the other species, except for the population at site 017, which overlaps with both *A. ndalalani* and *A. alcalica* populations on CVs 1-3 (Figure 5.16). Sampling site 017 is the site at which hybridisation between species was inferred from the genomic data (chapter three). Given the intraspecific separation seen from this analysis, separate analyses of species diversity were conducted as sub-analyses within species, which are presented in chapter six. Additional sampling sites (sites where genomic data was not available) were included for the within-species population-level analysis (see next chapter).

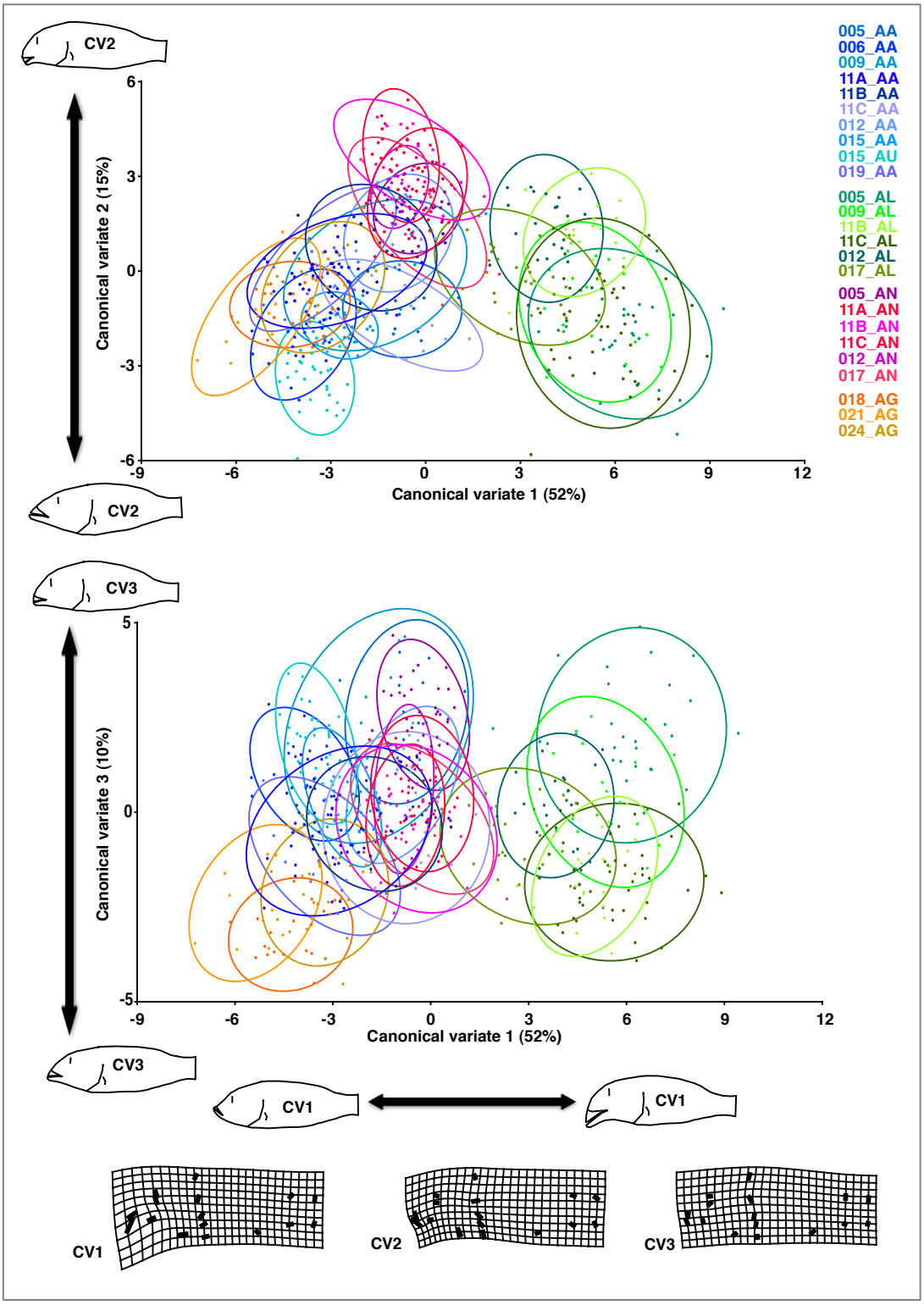


Figure 5.16. Canonical variate analysis of populations.

All species exhibit some degree of differentiation within species between sampling sites. Data is plotted for CV1 vs. CV2 and CV1 vs. CV3, with ellipses of equal frequency drawn at 90% probability. Outline drawings represent shape changes at the minimum and maximum extent of each CV.

Morphometric data are not considered reliable for inferring phylogeny (Klingenberg & Gidaszewski 2010), however can be used comparatively with molecular data to infer whether similar morphologies arose through convergence or common ancestry. Hierarchical cluster analysis on the Procrustes distances of population means generally separated the groups by species and location. Analysis included Ward's method of cluster analysis (which clusters consecutive groupings based on minimising within-group variance after merger; (Ward 1963)) and Neighbour-joining cluster analysis (which minimises total branch length at each stage of clustering; (Saitou & Nei 1987)). Both methods identified a clade containing *A. alcalica* and *A. grahami*, however the latter species nests deep within *A. alcalica* based on the NJ method, while these species are sister clades based on Ward's method (Figure 5.17).

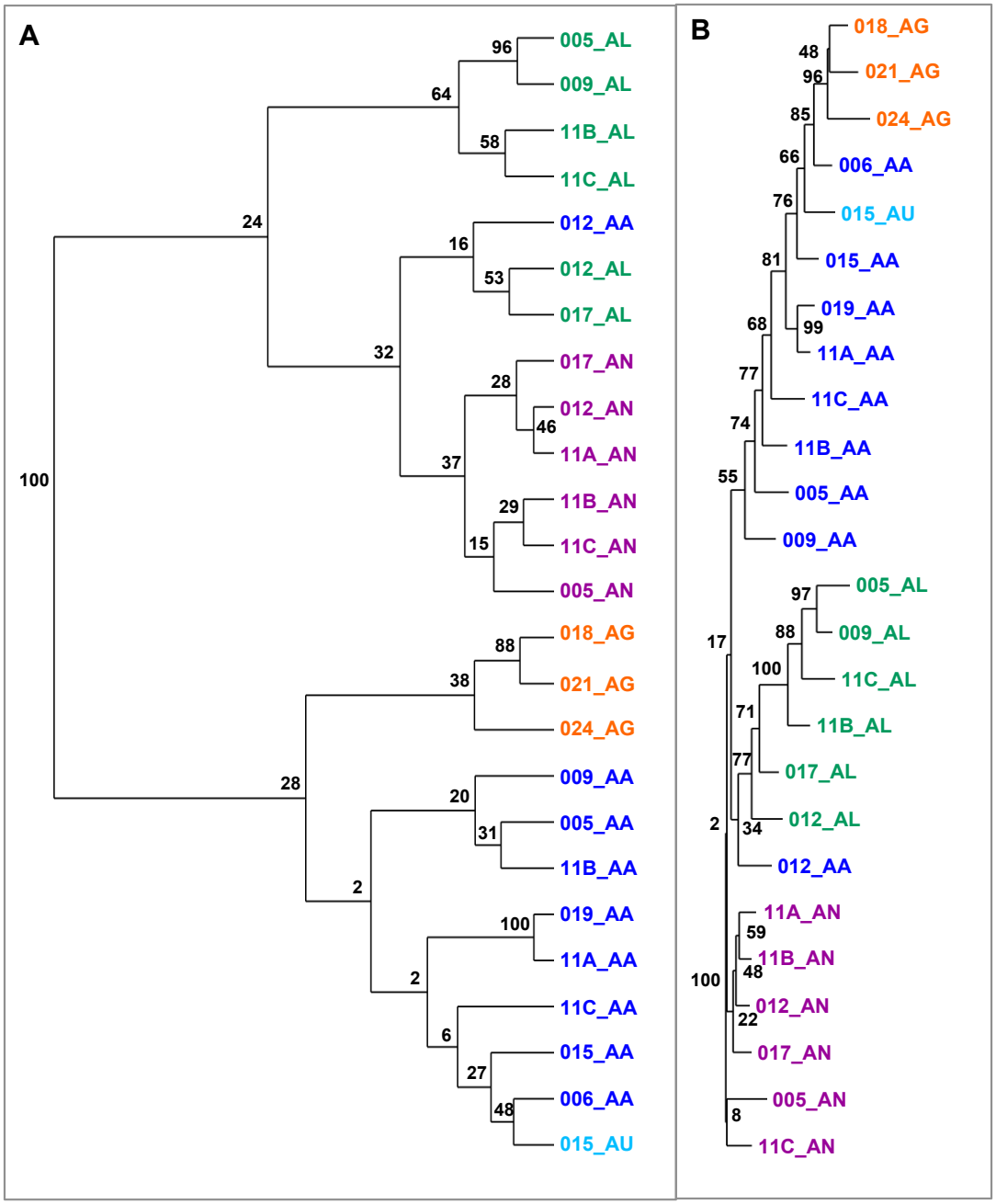


Figure 5.17. Phylograms from hierarchical cluster analysis based on geometric morphometric Procrustes distances of population means. Values on nodes indicate support for 100 bootstrap replicates. A) Ward's method B) Neighbour-joining method.

While the clustering methods generally reflect the correlation between morphology and phylogeny, the repeated clustering of site 12 *A. alcalica* population within *A. latilabris* in both analyses is unexpected, although support is very low. Reviewing the placement of these populations in morphospace indicates that site 12 *A. alcalica* is closest to site 17 *A. latilabris*, but only marginally more so than its proximity to other *A. alcalica* populations (Figure 5.18), highlighting the limitations of displaying 2-dimensional data as a bifurcating phylogram. It is of note that the population to

which site 12 *A. alcalica* is closest is the site 17 *A. latilabris* in PC1-PC2 axis and site 17 *A. ndalalani* in the PC1-PC3 axis, the sampling site that was considered to be of possible hybrid origin based on genomic analysis (chapter three).

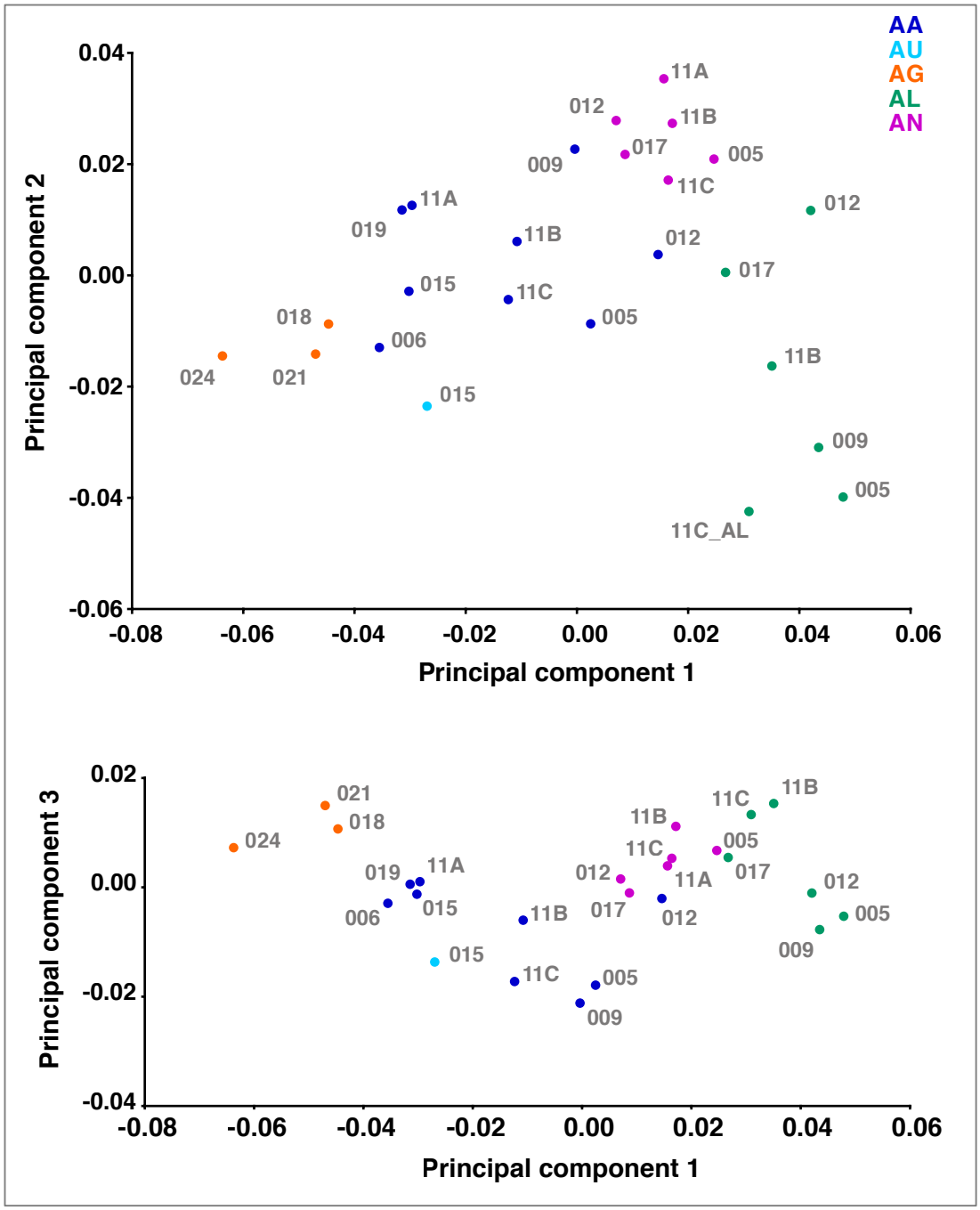


Figure 5.18. Labelled populations (sampling sites) for the PCA analysis.

Data shown are population means, coloured by species and with the grey label indicating sampling site. The population of *A. alcalica* from site 012 appears closest to *A. latilabris* (PC1-PC2) or *A. ndalalani* (PC-1-PC3).

Employing the same clustering methods at a species level clusters *A. latilabris* with *A. ndalalani* for both methods, with discrepancy in the placement of the southern clade of *A. alcalica*, which clusters with the northern *A. alcalica* clade using Ward's method, but with *A. latilabris* and *A. ndalalani* using Neighbour-joining (Figure 5.19). Perhaps more surprisingly, given the more upward mouth position of Lake Magadi populations, *A. grahami* clusters closest to the northern *A. alcalica* clade rather than the *A. alcalica* upturned-mouth morph.

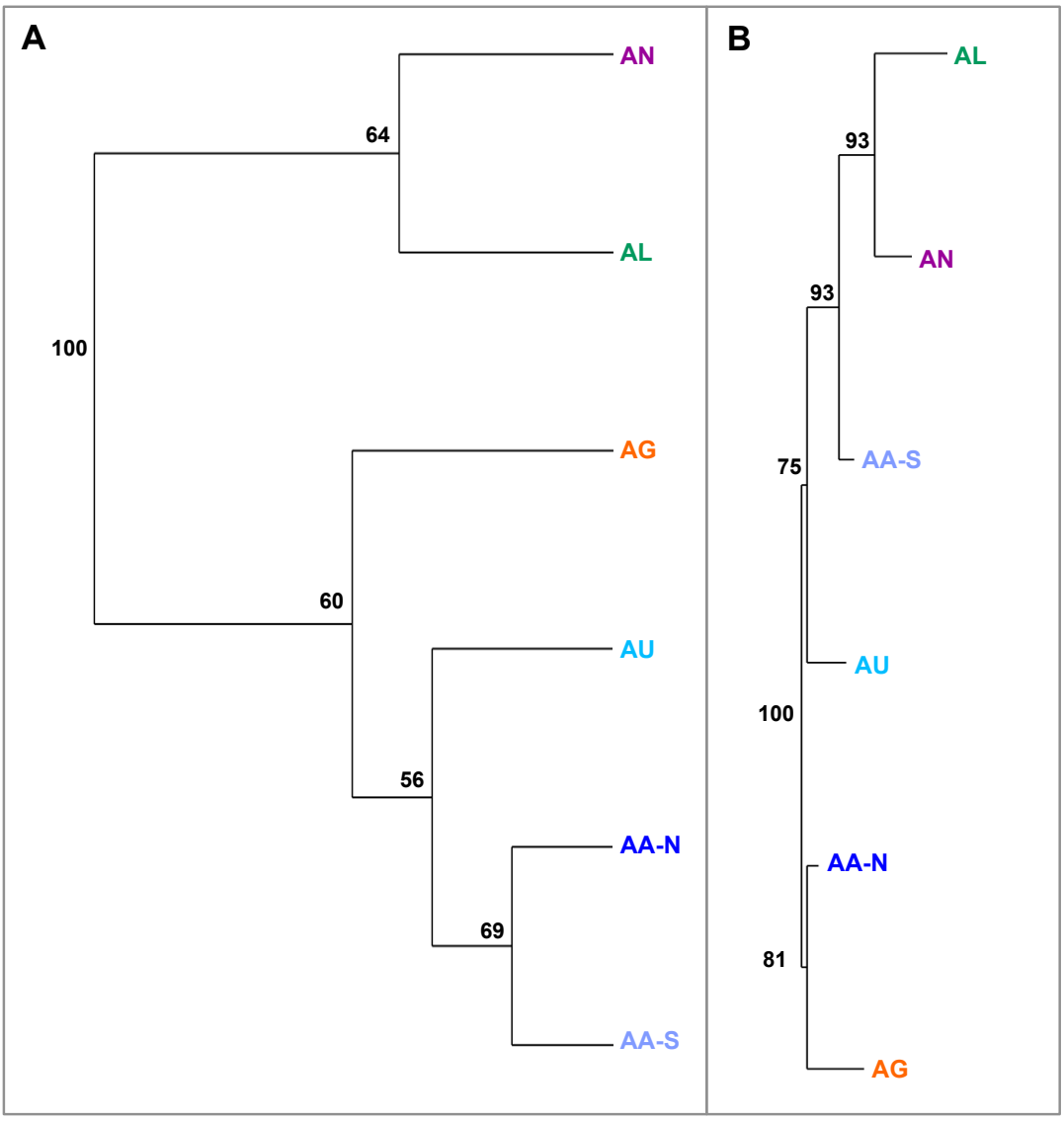


Figure 5.19. Phylograms from hierarchical cluster analysis based on Procrustes distances of species means. Values on nodes indicate support for 100 bootstrap replicates. A) Ward's method B) Neighbour-joining method.

Given the morphological differentiation of species and populations across the dataset and the clustering of species in phylograms based on morphological data,

we may expect a strong phylogenetic signal from the dataset. The effect of such a signal was tested using permutation analysis within MorphoJ based on the branch lengths of the molecular phylogeny from chapter three, and found to be significant. (unweighted square-change parsimony; permutation test against null hypothesis of no phylogenetic signal; 100,000 permutation; P<0.0001). Consequently, the effect of phylogeny was also tested for impact on PCA. These analyses used the reduced-taxon phylogeny from chapter three, and as such did not include the following sites: 12, 24, and site 11 subdivisions. As it was present in the phylogeny, site 17 *A. alcalica* was included in the comparison, but only for the *A. alcalica* individual that clustered with *A. alcalica* in the phylogeny and did not display an intermediate phenotype of *A. alcalica* / *A. ndalalani*. This population was not included in the other morphometric analyses of this chapter. As only one such individual was available for morphometric data, any inference from the results for site 017 *A. alcalica* should be treated with caution given the sample size here of n=1. Analyses were conducted in Mathematica: i) conducting a standard PCA with no correction and subsequently mapping the molecular tree onto the PCA, and ii) conducting a phylogenetic PCA, where the PCA is weighted by branch lengths. Both analyses showed very similar results, and despite the strong signal of phylogeny in the dataset, no apparent discrepancy is discernible between analyses (Figure 5.20). This lack of difference is presumably due to the fact that the distance in the genomic data is extremely shallow (chapter three), and so weighting by branch length has negligible effect on the large morphological differences demonstrated by the PCA.

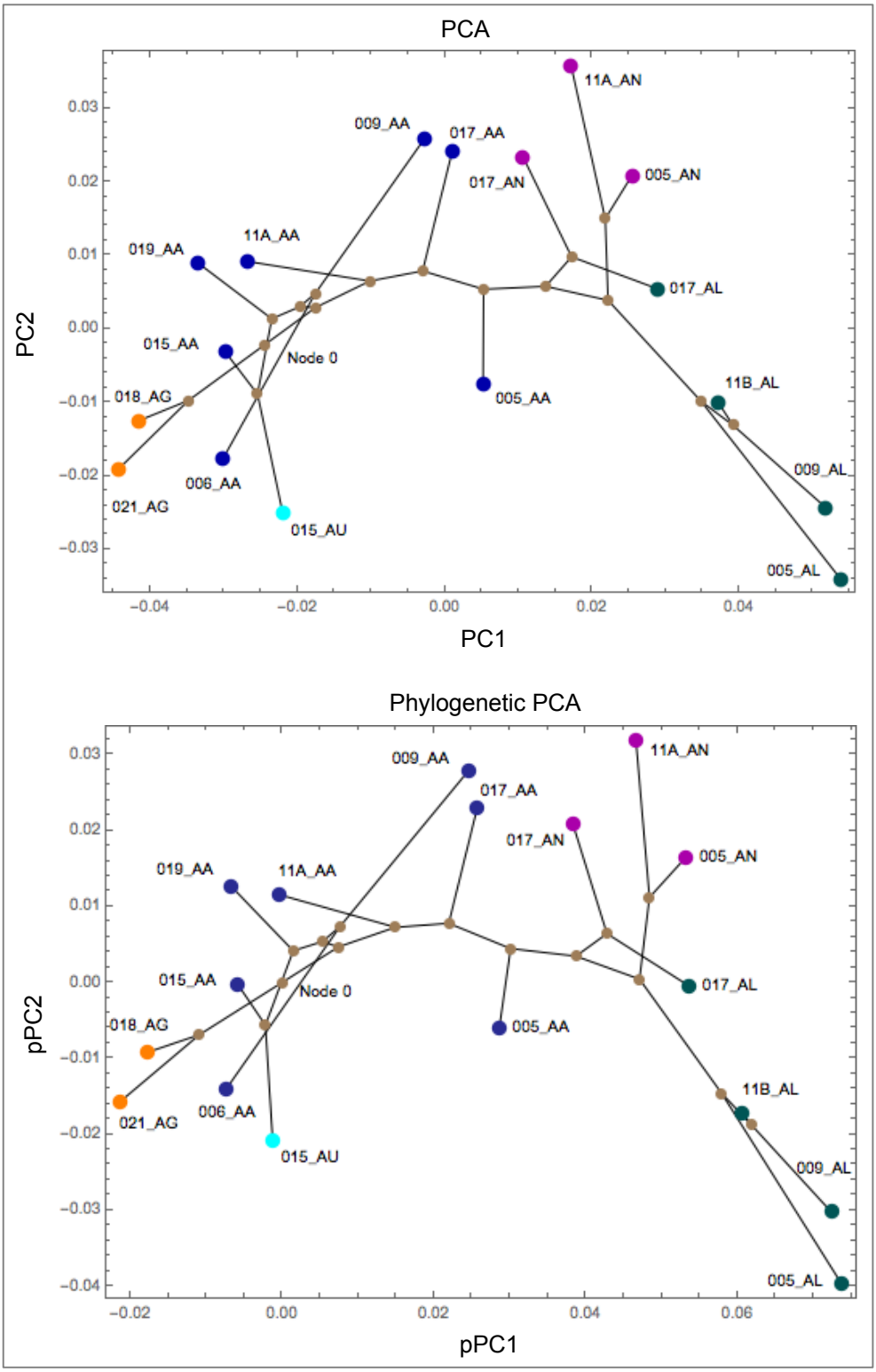


Figure 5.20. Comparison of standard PCA and phylogenetic PCA to assess the effect of phylogeny on morphological species relationships.

Mapping the ML phylogeny in the PCA morphospace with no correction (Figure 5.21) showed that there was some degree of discordance between genomic and morphometric differentiation, particularly due to site 17 individuals that clustered by species in morphospace, rather than by collection site as in the phylogeny. Although two of the northern *A. alcalica* populations (site 6 and site 15) are close to *A. grahami* in morphospace as might be expected from the genomic analysis, site 19 *A. alcalica* individuals appear closest to some of the southern populations (in particular, site 11).

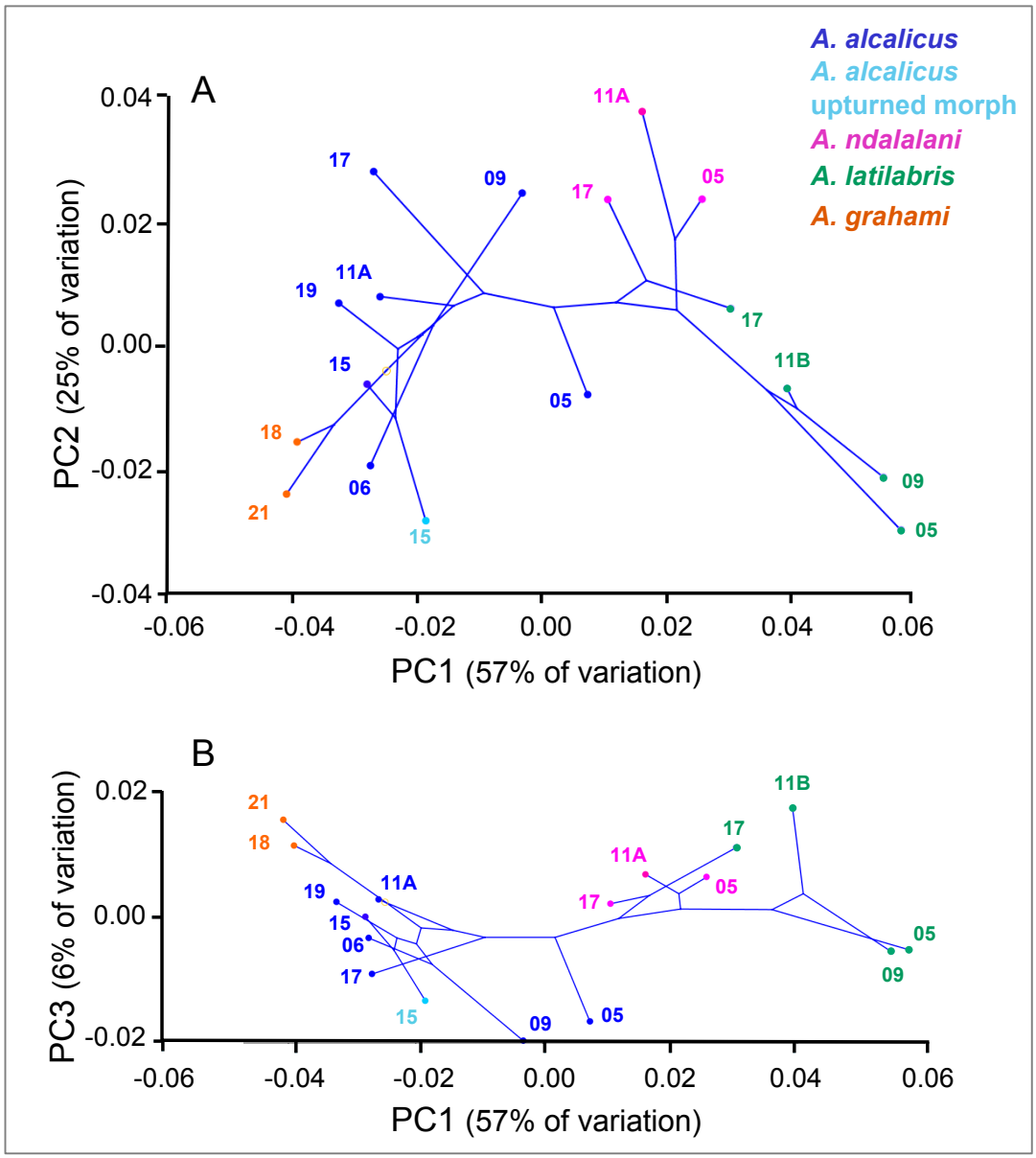


Figure 5.21. Phylomorphospace reconstruction of *Alcolapia*. RAD ML phylogeny (pruned to one individual per population) mapped in morphospace of A) PC1 vs. PC2 and B) PC1 vs. PC3 from the PCA across the full morphometric dataset. Numbers at terminals indicate sampling site (population).

The discordance between morphometric and genomic data is further demonstrated by the PCA of RAD data (Figure 5.22) where, in contrast to the body shape PCA, the Lake Magadi species *A. grahami* is distinctly separated from the Lake Natron species, although it should be noted that the PCs explain a substantially smaller amount of variation in this instance. The three Natron species are less tightly clustered, with overlap between all three species, and with *A. ndalalani* appearing intermediate between *A. latilabris* and *A. alcalica*.

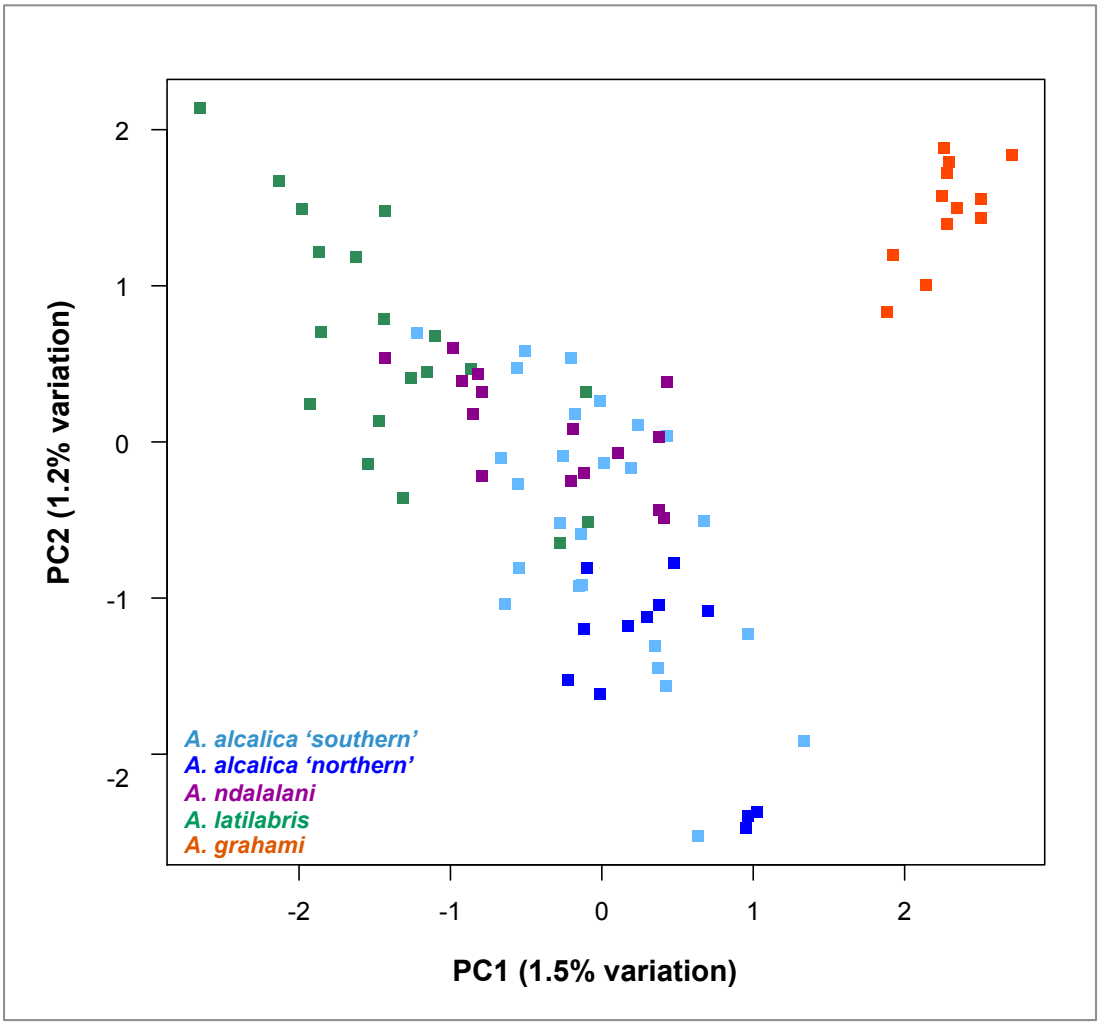


Figure 5.22. Principal components analysis of genomic variation in the RAD dataset using unlinked sites (818 SNPs).

Given this discordance of genomic and morphological variation within the dataset, morphological (Procrustes) distance was plotted against phylogenetic distance from the molecular phylogeny to assess variation in comparative change. An arbitrary value of 0.5 was used as a ratio threshold to consider those pairwise comparisons

that displayed non-correlation of phylogenetic and morphological distance (i.e., those comparisons falling outside of y=0.5x and y=2x) to indicate comparisons outside the common mode, however these were not tested statistically (c.f. (Muschick *et al.* 2012, Figure 4). These thresholds are indicated by grey shading in Figure 5.23. None of the comparisons fell above this axis on the morphological axis indicating rapid divergence relative to the rest of the dataset, however it should be noted that this is because comparisons are scaled within the dataset, and if absolute values were plotted, all comparisons would indicate rapid morphological divergence compared to phylogenetic distance (plotted for comparison in Figure 5.24). However, there were several comparisons that showed comparatively greater phylogenetic distance than morphological distance, indicating stasis or convergence in body shape (e.g., Muschick *et al.* 2012), which were mainly intraspecific (i.e., population) comparisons but also included several comparisons of *A. alcalica* and *A. grahami*, as well as two comparisons between *A. alcalica* and *A. ndalalani* populations.

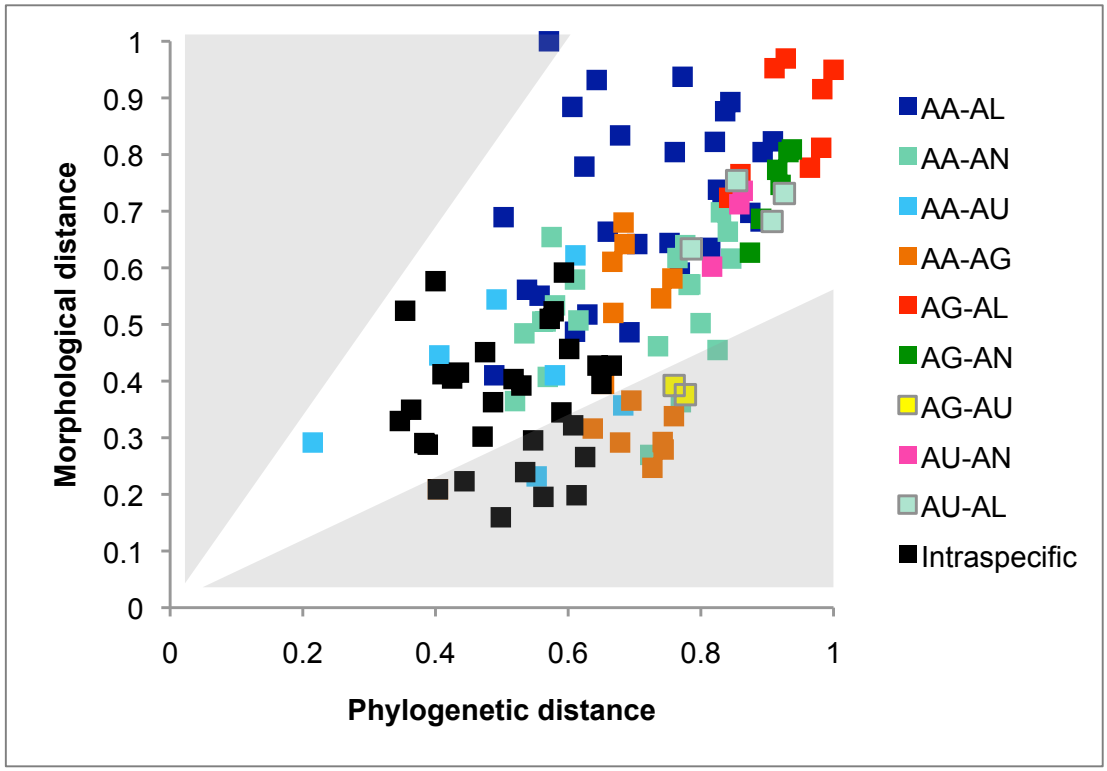


Figure 5.23. Phylogenetic distance plotted against morphological distance.
Population pairwise distances are scaled to 1. Grey shading indicates regions of expected comparative rapid divergence (top left) or convergence (bottom right) based on arbitrary thresholds of y=0.5x and y=2x. Intraspecific designation indicates population comparisons within each of the described species.

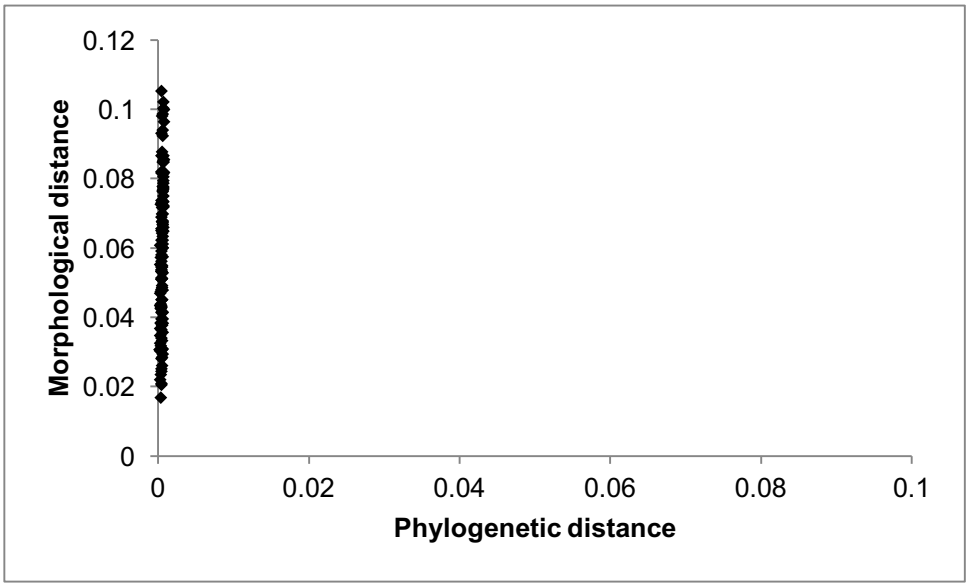


Figure 5.24. Absolute values (non scaled) of phylogenetic distance plotted against morphological distance.

In order to investigate whether these latter cases may be examples of stasis or convergence, ancestral state reconstruction of internal nodes of the ML tree was conducted in MorphoJ using weighted squared-change parsimony. The reconstructions suggest a common ancestor with terminal-to-moderately upturned mouth features, less marked mouth orientation than the more upturned descendent *A. alcalica* and *A. grahami*, suggesting morphological convergence or stasis in comparison to the highly divergent sub-terminal mouth forms *A. ndalalani* and *A. latilabris* (Figure 5.25).

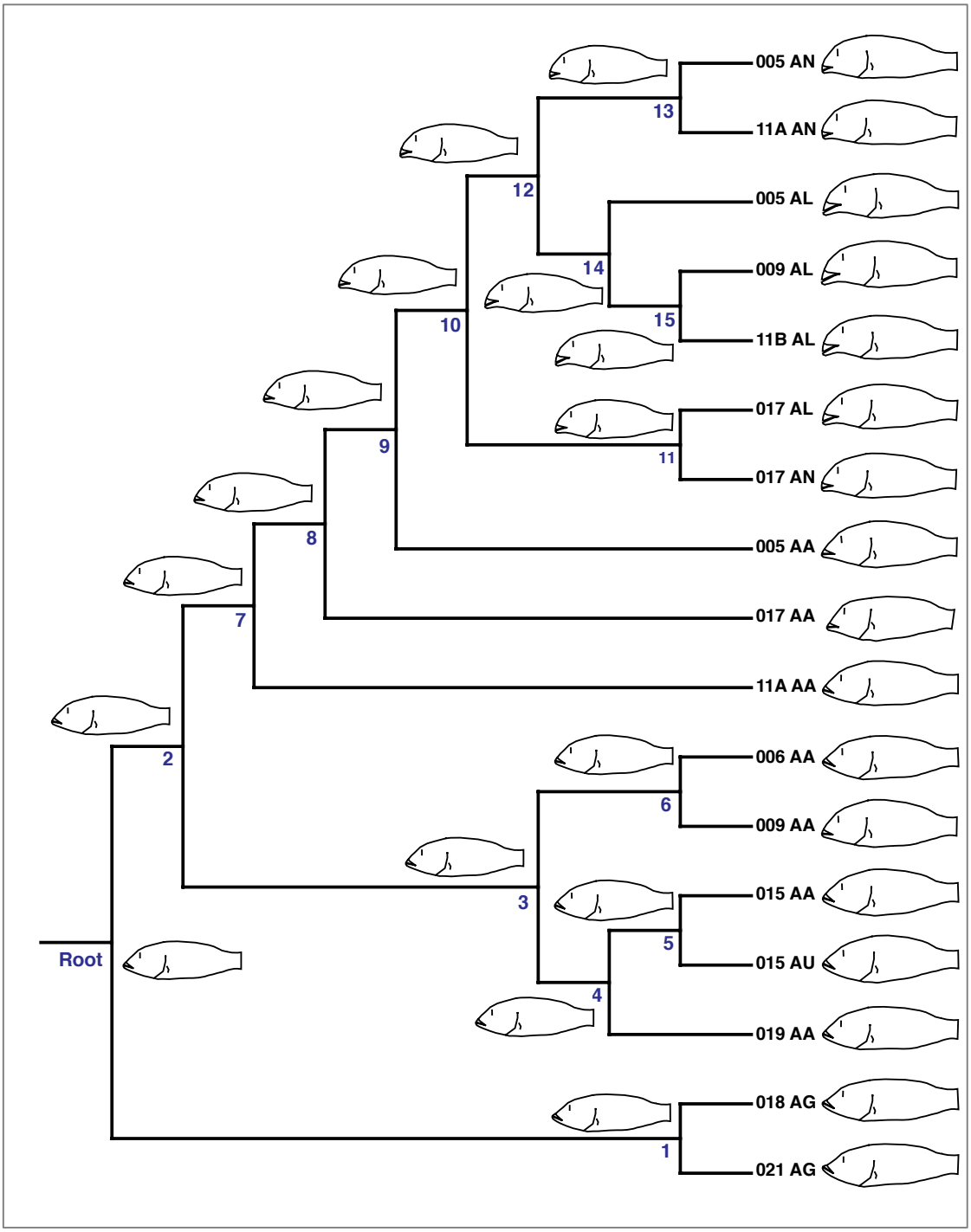


Figure 5.25. Ancestral shape reconstruction of body shape across the molecular phylogeny. Reconstructed shapes for internal nodes are placed above the respective nodes – except for four nodes for which the shapes are placed below (nodes 4, 14, 15) or to the right (root) of the respective node due to space constraints. Shapes at tips indicate consensus shape for populations in the molecular phylogeny.

Convergence was explicitly tested using the SURFACE package in R. During the forward phase of analysis k=3 regime shifts were detected (AIC improvements from -145.2477 to -157.1446), and these were not collapsed further during the backwards

phase. No shifts were found towards a convergent phase (c=0), and all regime shifts were designated as non-convergent evolution. Visualising these regime shifts on the tree suggested that adaptive divergence corresponded to an ancestral *A. alcalica* / *A. grahami* morphotype, with an adaptive shift to inferior mouth morphology (*A. ndalalani* and *A. latilabris*) and a second shift to hypertrophied lips (*A. latilabris*) (Figure 5.26). Notably, one of the southern *A. alcalica* populations (site 5) appears to undergo the first regime shift with the ancestor of *A. ndalalani* and *A. latilabris*.

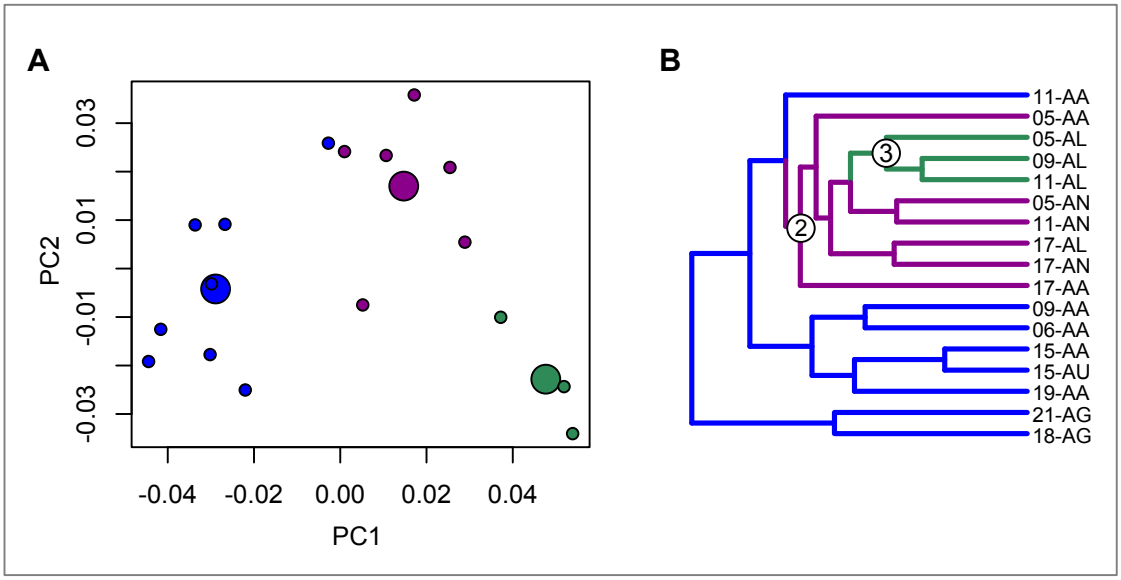


Figure 5.26. SURFACE analysis to detect convergence.

A) Plot of PC1 (55%) vs. PC2 (23%), with trait values for each population as small circles and trait optima for each regime shift as large circles. The dispersion of trait values around trait optima (i.e., not tightly clustered at optima) indicates morphological variation within each regime. B) Cladogram with regime shifts painted on branches. Numbers on the branches indicate the order in which the regime shifts were added during the forward phase. None of the regime shifts exhibited convergent evolution, with all adaptive divergence indicating non-convergent evolution.

Re-running the SURFACE analysis without the site 17 *A. alcalica* population, which had low sample number (n=1), produced the same results and number of regime shifts, but removing other site 17 populations (for *A. latilabris* and *A. ndalalani*) reduced the number of regime shifts to two (Appendix Figure 5A.3).

Body shape variation between populations

Examples of preserved specimens of *Alcolapia* species and morphs from different populations sampled in the present study are shown in Figure 5.27. The *A. alcalica*

specimen depicted from site 17 is a confident species ID (i.e., it is not one of the specimens reassigned to *A.* aff *ndalalani* following the genomic analysis). Photographs of all *A. alcalica* specimens from site 17 are shown in Appendix Figure 5A.1. From these photographs, it can be seen that the distinctive features for each species exhibit variation, for example site 17 *A. latilabris* has a sub-terminal mouth and thickened lips but to a lesser degree than individuals from sites 5 and 11, while the characteristic blunt snout of *A. ndalalani* from site 17 is less so than in specimens from sites 5 and 11. This variation in morphology may reflect the genomic analysis (chapter three) indicating that rates of interspecific gene flow are highest at site 17. The intraspecific variability of morphology is further discussed in chapter six.

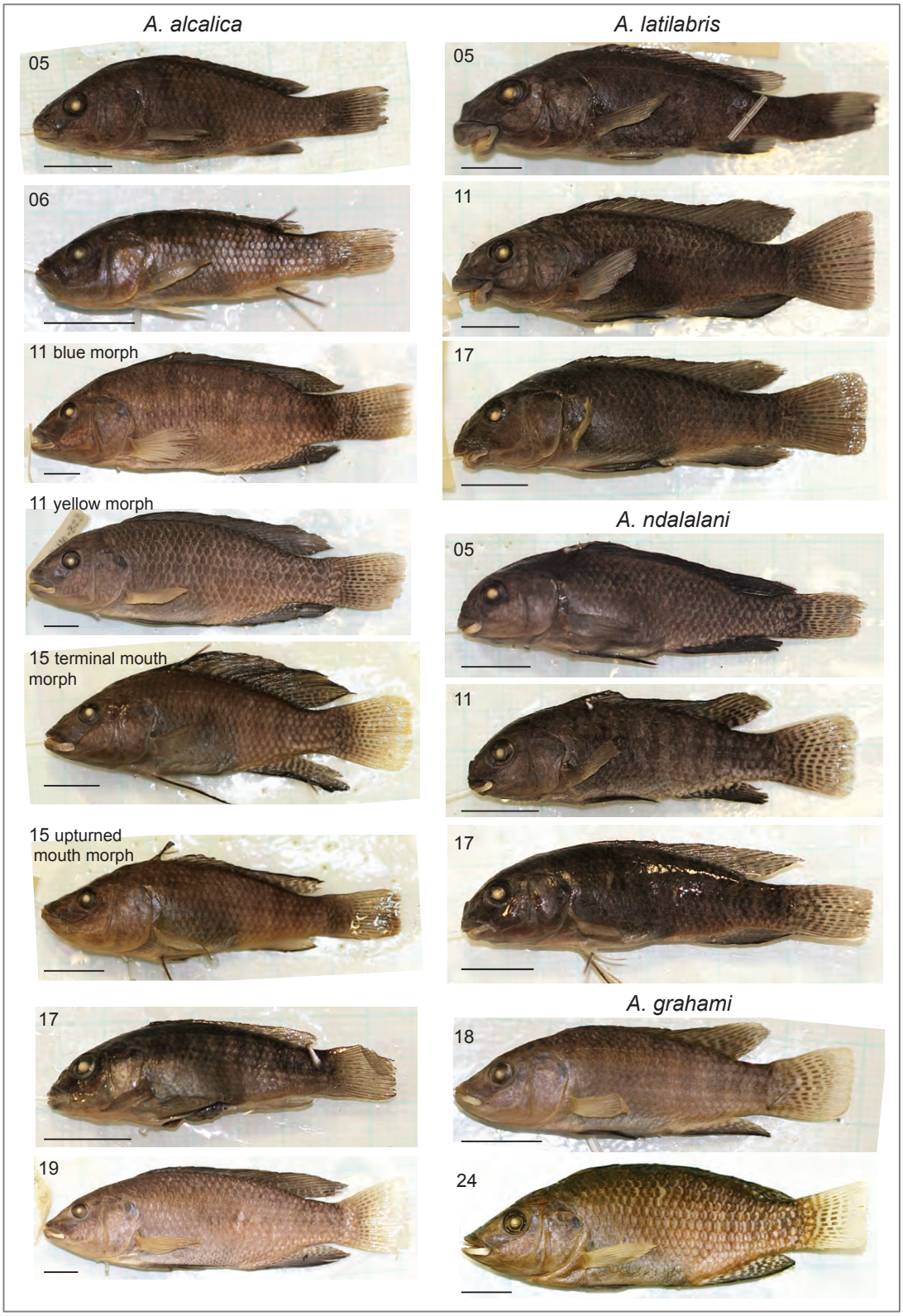


Figure 5.27. Preserved specimens of *Alcolapia* species and morphs.

All photographs are of male specimens. Number in top left hand corner of each photograph is the sampling site for each specimen. Black bars: 10mm. Further information on sampling site locations is given in Figure 5.1. Colour photographs of live specimens are shown in Figures 1.4 and 3.1.

Variation in lower PHJ bone shape

A total of 108 individuals across five sampling sites were included for the geometric morphometric analysis of the lower pharyngeal jaw (PHJ) bone. Regression of Procrustes coordinates against centroid size was significant (10,000 permutations; $P<0.0001$) and size accounted for 6.80% of the variation within the dataset, which was similar to the proportion of variation explained by size in the body shape data (5.6%). In the PCA on the Procrustes-fitted alignment, the first three PCs accounted for 77% of the variation within the dataset (all other PC Eigenvalues accounted for <9% variation each; total of 10 PCs; Appendix Table 5A.2).

Species were minimally separated by PCA, but clustered by species in CVA (Figure 5.28) although to a lesser extent than that shown in the body shape data. Shape changes for both analyses described a narrowing and lengthening of the lower PHJ bone (PC1) or a flatter, broader tooth surface (PC2). In CVA, all pairwise comparisons were significant after Bonferroni correction ($P<0.05$), except for *A. grahami* compared to *A. alcalica* upturned morphs (Table 5.5). In non-parametric comparisons (NPMANOVA), all pairwise comparisons were significant except those involving *A. grahami* and the upturned mouth morph of *A. alcalica* (Table 5.6), which may be explained by the smaller sample size for these groups.

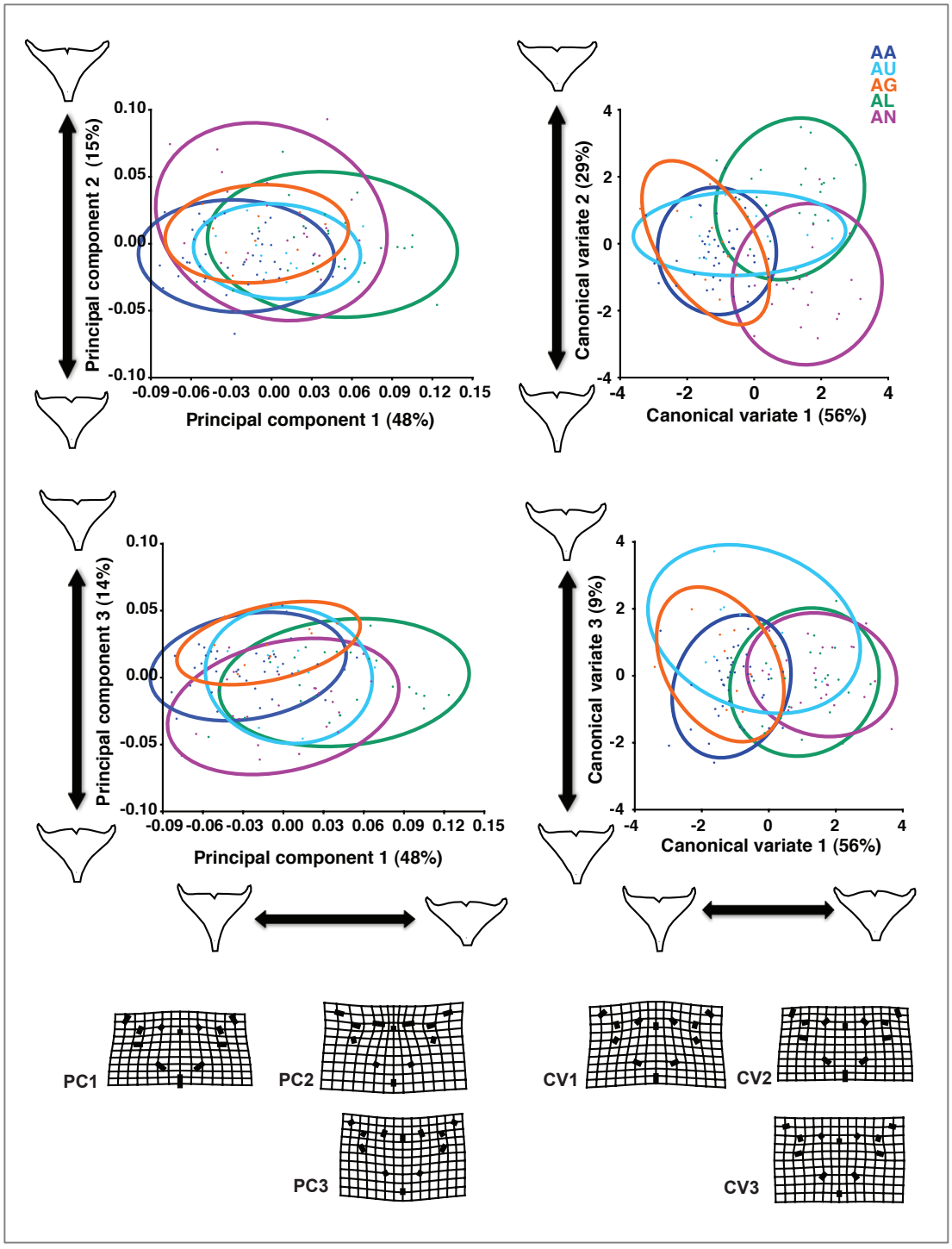


Figure 5.28. PCA and CVA of PHJ shape variation between species. Outline shape drawings represent the shape at the minimum and maximum extent of data along each PC axis. Warped transformation grids show maximum change from consensus shape along the positive axis only.

Table 5.5. Interspecies-distances from canonical variate analysis of lower pharyngeal jaw shape, all individuals.

	AA	AG	AL	AN	AU
AA (n=41)	-	0.03*	0.07*	0.05*	0.04*
AG (n=9)	1.60*	-	0.07*	0.05*	0.03
AL (n=25)	2.69*	3.11*	-	0.05*	0.05*
AN (n=23)	2.90*	3.47*	2.61*	-	0.04*
AU (n=10)	1.93*	2.2*	2.47*	2.92*	-

*Pairwise comparisons significant at P < 0.05 (10,000 permutation rounds).

Below diagonal: Mahalanobis distances among groups;

Above diagonal: Procrustes distances.

Table 5.6. Pairwise F-values for NPMANOVA between species for lower pharyngeal jaw data.

	AA	AG	AL	AN	AU
AA (n=41)	-	2.67	28.2*	9.97*	3.76
AG (n=9)	1.84	-	8.69*	4.72*	2.194
AL (n=25)	3.84*	2.15*	-	8.74*	4.52
AN (n=23)	3.86*	1.94*	2.22*	-	2.34
AU (n=10)	2.13*	1.46	1.87	1.58	-

*Significant at P < 0.05 for Bonferroni-corrected P-values (10,000 permutations).

Below diagonal: Mahalanobis distances among groups;

Above diagonal: Euclidean distances.

The sampling design for the pharyngeal jaw data was originally developed to consider species differences, as well as inter-population differences within species (between sites). However, as this resulted in varying sample sizes between species, analysis was also conducted on a subset of data including only Lake Natron species occurring sympatrically (at sites 05 and 12) to ensure even sample numbers for each species. This subset included a total of 70 individuals (*A. alcalica*: n=22; *A. latilabris*: n=25; *A. ndalalani*: n=23), and regression analysis revealed that size accounted for 7.5% of the total variation (P=0.02). Eigenvalues for PCA of this data subset are given in Appendix Table 5A.3. Loadings from the PCA were similar to that for the larger PHJ dataset, with 80% of variation explained by the first three variables (Appendix Table 5A.3). The PCA showed similar differentiation to the full dataset (Figure 5.29). Examination of variation within species between sampling sites is considered further in chapter six.

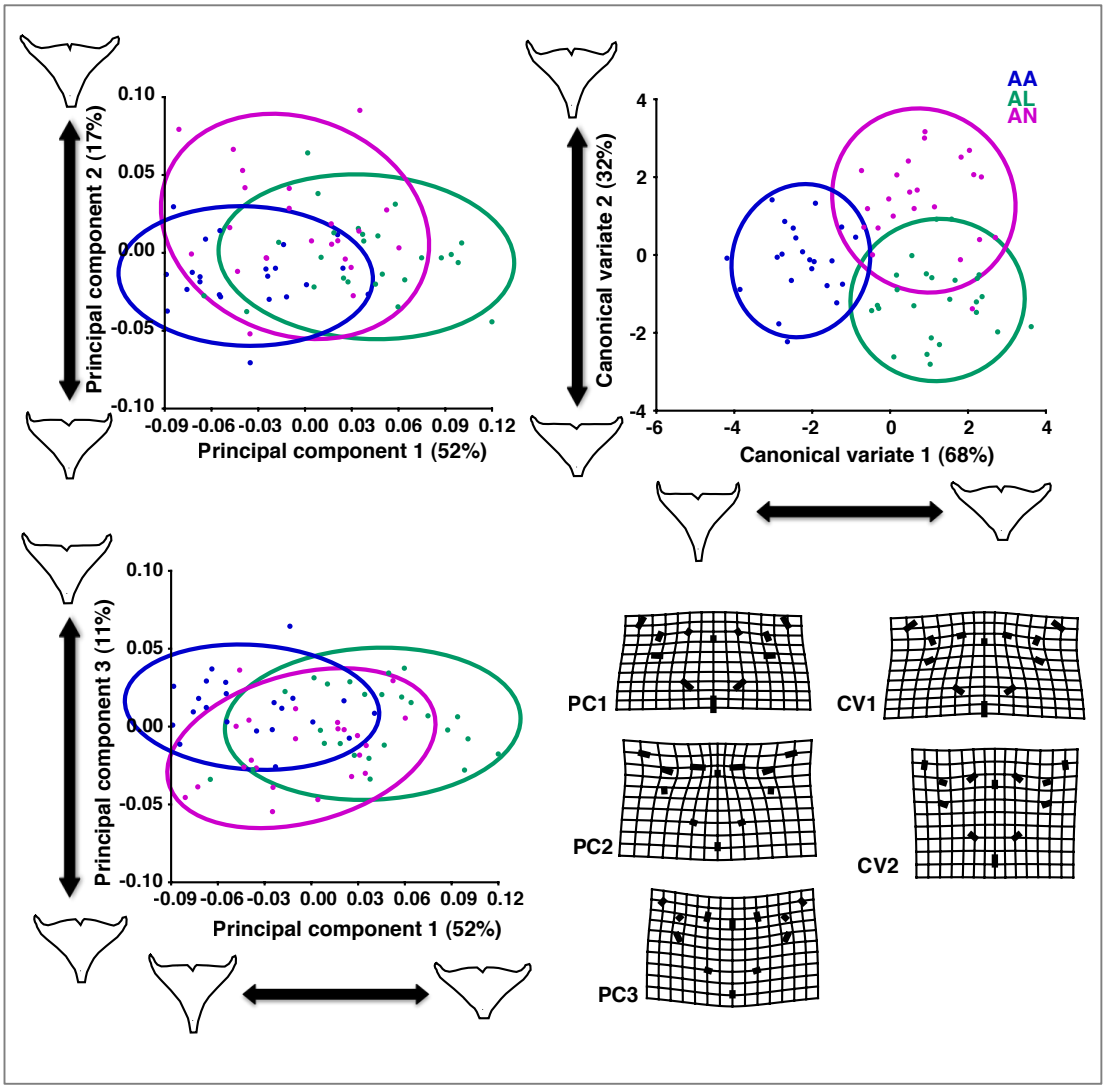


Figure 5.29. Shape analysis of lower pharyngeal jaw for sympatric populations.
Two CVs are produced for this comparison as only three groups are included.

Across-site comparisons and dataset covariation

Morphometric and dietary covariation

The partial least squares regression analysis in MorphoJ revealed that $\delta^{13}C$ varied with morphology (correlation: 0.44453; P-value from 10,000 round permutation test: P<0.0001) but $\delta^{15}N$ did not (correlation: 0.19914; P-value from 10,000 round permutation test: P=0.2814). This reflects that the SIA $\delta^{13}C$ values differ significantly between species, and the species are significantly differentiated in morphospace.

Dietary and phylogenetic covariation

Stable isotope values also co-varied with phylogenetic distance in all sympatric sites excluding site 12 (Mantel test on inter-specimen RAD p-distance vs. SIA Cartesian

coordinate distance; P<0.05; Table 5.7). The lack of correlation in the site 12 individuals may be due to lower statistical power (fewer individuals for site 12 RAD data, as multiple individuals from this site were excluded from all analysis due to low sequencing quality) or that site 12 individuals do not form a distinct clade in the ML phylogeny.

Table 5.7. Mantel test results for RAD interspecimen p-distance vs. interspecimen SIA (Cartesian) distance (sites where multiple species present).

Site	Observation	P-value
Site 005	0.38	0.02
Site 009	0.41	0.03
Site 011	0.55	0.0001
Site 012	0.09	0.25
Site 017	0.34	0.03

Morphometric and phylogenomic covariation with geography

Between-population intraspecific morphological distances (Mahalanobis and Procrustes distances) for covariation tests were calculated by canonical variate analysis in MorphoJ. Plots of these CVA analyses for RAD-sequenced populations are shown in Appendix 5 (Figure 5A.4), but not discussed further here. Investigation of morphological differentiation and specific shape changes between populations is discussed in chapter six.

Simple Mantel tests revealed no significant correlation of morphometric distance with geographic distance or genetic distance. Partial Mantel tests for the association between genetic divergence and adaptive divergence controlling for geographic distance (testing for isolation by adaptation) revealed a significant correlation only for *A. ndalalani*, suggesting that populations that are more morphologically differentiated are also more genetically diverged (Appendix Table 5A.4).

Discussion

Despite *Alcolapia* being the focus of several morphological studies and descriptions, especially due to the large extent of morphological variation within a small species flock, this differentiation has not previously been investigated in an ecological context. The large differences in head and trophic morphology have been suggested to represent divergence in trophic resources or feeding strategies, however such ecological differentiation has not previously been tested. Here, morphometric analyses are used to define and quantify morphological variation within the group, and paired with SIA, gut length and stomach contents analyses to examine ecological variation. Implementing a phylogenetic framework, these analyses demonstrate ecological niche partitioning within the *Alcolapia* species flock for the first time.

Trophic niche space differentiation as an indicator of ecological speciation

Analysis of trophic niche space (using stable isotope ratios as a proxy for dietary resource utilisation) demonstrated significant differentiation of *A. alcalica* from *A. latilabris* and *A. ndalalani* at all sites where they occurred sympatrically with the exception of site 17 (Figure 5.4). Differentiation was observed only in the $\delta^{13}C$ isotope ratio, which indicates primary carbon source, and not in $\delta^{15}N$ ratio, which serves as an indicator for trophic level. As expected given previous stomach content analyses (Coe 1966) and anecdotal feeding observations (Albrecht *et al.* 1968; Trewavas 1983), the Lake Natron species all appear to be feeding on vegetation at the same trophic level. However, *A. alcalica* is identified as feeding on a different carbon source. These findings reflect the differences in trophic morphology, with *A. alcalica* having a terminal mouth (suggestive of surface and water column feeding on free-floating algal matter) and the other species both having sub-terminal mouths, more adapted to benthic feeding and algal scraping of attached resources; a behaviour that was observed in *A. latilabris* during fieldwork (pers. obs.). Divergence along a similar benthic axis (trophic resource utilisation, feeding on attached vs. unattached prey/matter) has previously been implicated in other East African cichlids, in pairs of ecotypes of the Lake Victoria genus *Neochromis* (Magalhaes *et al.* 2012) and Lake Barombi Mbo *Pungu* (Schliewen & Klee 2004).

Reviewing these results in line with the hypothesis that *A. alcalica* is surface-feeding, planktonic algae is generally $\delta^{13}C$-depleted relative to benthic algae (Hamilton & Lewis 1992; France 1995; Solomon *et al.* 2011), whereas in the present study *A. alcalica* is consistently $\delta^{13}C$-enriched relative to the other two species at all sympatric sites. However, producer isotope traces vary widely between different environments (Hamilton & Lewis 1992), and it is not possible to infer the exact dietary resources utilised without corresponding baseline samples. As such, the complementary analysis of stomach contents is useful to disentangle these trophic niche patterns. Stomach contents reflected SIA results, where *A. alcalica* was differentiated from other species while *A. latilabris* and *A. ndalalani* overlapped in diet (Figure 5.6). Stomach contents of *A. alcalica* showed a high proportion of cellulose (Figure 5.5), and as vascular plants are not present in the soda lakes, it suggests that *A. alcalica* is feeding on terrestrial-derived plant matter. These sources may either be wind-blown or washed onto the surface of the springs, or could be from access to terrestrial sources if water levels rise during floods (e.g., Jackson *et al.* 2012). Furthermore, the stomach contents data support further differentiation of the benthic specialists by picking (*A. ndalalani*) vs. scraping (*A. latilabris*) foraging modes, based on the observation that *A. latilabris* stomach contents a greater proportion of grit/sand and grit of a greater particle size relative to *A. ndalalani* (Figure 5.5).

The overlap in niche space at site 17 could be due to low statistical power from insufficient sampling of *A. alcalica* at this site (n=4), or may reflect the results from the genomic data (chapter three), that the three sympatric species are not well distinguished at this site because of possible hybridisation, and therefore may not be exploiting separate ecological niches. Different trophic niche utilisation compared to other populations is observed at this site, indicated by patterns of niche breadth, with *A. latilabris* and *A. ndalalani* occupying considerably larger niche space at this site than any other sampled site (Figures 5.4). Furthermore, differential diet analysis in other freshwater systems has shown mate preference using SIA, suggesting overlapping trophic niches could indicate incomplete reproductive isolation between species (Snowberg & Bolnick 2012).

Morphological divergence in a young species flock

Geometric morphometric analysis separated individuals by species and population, with less intraspecific variation between *A. ndalalani* populations than between populations of other species (Figures 5.10, 5.11, Tables 5.2, 5.3). Between species,

PC1 of the PCA, which described differentiation based mainly on lip size, mouth orientation and body depth, separated *A. latilabris* and *A. ndalalani* from the other three species. Meanwhile PC2, which described differentiation in snout length and head length, separated *A. latilabris* from *A. ndalalani*, although there was a large degree of variation within *A. latilabris* populations across PC2. It should be noted that one of the main diagnostic features between these species is the fleshy lips of *A. latilabris*, and the landmark configuration did not account for lip area or width. Future analyses could consider the variation of lip size within and between species (e.g., Manousaki *et al.* 2013). *Alcolapia grahami* largely overlaps with *A. alcalica* on PC2 (though to the lower extent of the species distribution), but is differentiated by PC3 (Figure 5.10, 5.11). These findings are in accordance with findings of morphological differentiation between Lake Natron species using traditional morphological measurements in previous studies (Zaccara et al. 2014), but the additional inclusion of Lake Magadi *A. grahami* for landmark analysis in this study demonstrates the overlap in morphospace of *A. alcalica* and *A. grahami*, a morphological similarity noted by Tichy and Seegers (1999). Furthermore, the upturned-mouth *A. alcalica* morph at site 15 clusters most closely with *A. grahami* in morphology, but with *A. alcalica* in phylogeny, suggesting that the similar mouth morphology has arisen independently in these two forms (although not indicated by SURFACE analysis). More detailed analysis of the site 15 morphotypes is included in chapter six.

The clustering algorithms on morphological data consistently separated the Lake Natron species, but clustered northern *A. alcalica* populations (sites 6, 9, 15) with *A. grahami* (Figure 5.18, 5.19). Despite the clear signal of phylogeny on morphology, the application of phylogenetic PCA through weighting of branch lengths did not affect the outcome of the PCA comparison (Figure 5.20). Thus, the discordance between morphology and phylogeny for *A. grahami* and *A. alcalica* relationships is clearly seen when the phylogeny is mapped in morphospace (Figure 5.21), and highlighted by different patterns of differentiation in the PCA of genomic data (Figure 5.22) relative to those on body shape (Figures 5.10,5.11), as well as the comparison of morphological and phylogenetic distance (Figure 5.23). This similarity of mouth orientation is seen most clearly through the reconstruction of ancestral body shape (Figure 5.25). The reconstruction suggests that the *Alcolapia* species flock ancestor had a terminal to moderately upturned mouth, which remained as terminal in the southern populations of *A. alcalica* but diverged to varying levels of upturned orientation in northern Natron *A. alcalica* and *A. grahami* populations.

The morphological differentiation seen in lower pharyngeal jaw shape is markedly less pronounced than in the overall body shape (Figure 5.28, 5.29). While the species are still statistically differentiated by the PHJ data, there is a large amount of overlap across all the species. These patterns correlate with differentiation of the stable isotope ratios observed in the present study. Such correlation is similar to findings in the Lake Tanganyika cichlid radiation, where differences in body shape correlated with differences in carbon source ($\delta^{13}C$), while differentiation in PHJ shape was reflected by differences in trophic level ($\delta^{15}N$) (Muschick *et al.* 2012). As most variation in PHJ shape is correlated with food hardness rather than food type, it is perhaps not surprising that there is a lower extent of variation within PHJ shape than body shape, as all *Alcolapia* species are mostly herbivorous. Despite lower variability, differentiation is seen between all Lake Natron species PHJ shape in the CVA analysis, while *A. alcalica* and *A. grahami* are not significantly differentiated by PHJ shape, suggesting very similar food processing habits. Unfortunately, the small size of the specimens meant that several PHJ jaw bones were extracted without the full lower process, so it was not possible to include a landmark at the very tip of the bone, which may have increased the power of the analysis to separate species, given that the main differentiation along CV1 axis was describing relative height/width (Figures 5.28, 5.29). Furthermore, the inclusion of additional specimens per species would enable the use of additional semi-landmarks in describing the bone outline. Although not considered here, analysis of tooth shape (both for pharyngeal and oral jaws) may further differentiate species and inform on foraging mode, as differences in cusping and tooth arrangement have previously been described between these species (Tichy & Seegers 1999).

That the species are separated by body shape morphometrics is possibly not surprising, given that their original descriptions were based on external morphology (Table 1.3), but the current chapter demonstrates intraspecific differentiation between certain populations (Figure 5.21). Two of the most northerly *A. alcalica* populations (site 006 and site 015) cluster with *A. grahami*, but the most northern population (site 019) does not, which is incongruent with the genomic results, although it is possible that there is a habitat effect as site 006 was considerably more shallow than other Lake Natron sites (~5 cm) and more similar to Magadi sites, while site 019 was at the edge of a freshwater river, and considerable cooler and deeper than Magadi sites. In both cluster analyses and the PCA (Figures 5.17-5.19) the *A. grahami* population from Lake Nakuru (site 024) has the longest branch and greatest morphological distance from other populations, indicating the body

shape differences that have previously been noted in this population (Vareschi 1979; Trewavas 1983).

Plotting morphological distance against genetic distance (Figure 5.22) suggested that *A. alcalica*/*A. grahami* comparisons may represent stasis or convergence. Reconstruction of ancestral shape (Figure 5.25) suggested a terminal-mouthed ancestor, with mouth orientation becoming comparatively upturned in *A. grahami* and northern *A. alcalica* populations. SURFACE analysis supported these results (Figure 5.26), indicating that *A. grahami* and *A. alcalica* represent the ancestral body shape, with *A. ndalalani* and *A. latilabris* experiencing regime shifts and diversification.

Ecological niche partitioning in *Alcolapia ndalalani* and *Alcolapia latilabris*

The lack of differentiation in trophic niche between *A. ndalalani* and *A. latilabris* (Figure 5.4) is perhaps surprising given the large differentiation in trophic morphology (Figures 5.10, 5.10, 5.27). Although both species have sub-terminal mouths, which could indicate a similar benthic feeding behaviour, differences in lip size and snout length are pronounced. Thick-lipped species (similar to *A. latilabris*) have evolved multiple times in cichlid radiations, and have generally been associated with a diet of hard-shelled invertebrates and crustaceans and for feeding in rocky environments and between crevices, although the exact function is not yet known (Salzburger 2009). Recently, it has been suggested that thick-lipped cichlid morphs tend to have correspondingly narrow or pointed heads, and this feature enables foraging in crevices and rocks (Baumgarten *et al.* 2015). Frequently, other thick-lipped species/morphs have terminal or superior mouths, suggesting that the *A. latilabris* sub-terminal/inferior fleshy lips may provide a different trophic function. Furthermore, *A. latilabris* does not appear to have a narrower head than other species in the *Alcolapia* flock, although, morphometric analysis was not conducted dorsally, so this has not been tested. Certainly, *A. latilabris* does not feed on invertebrates, as indicated by the trophic level inferred from stable isotopes indicative of feeing on primary producers. It was apparent that the soda springs contain an impoverished aquatic invertebrate fauna (pers. obs.), although invertebrate traps were not set during fieldwork. One possibility is that the innovation arose during periods of higher lake levels when waters were less saline and more ecological niches were available (supporting a greater range of invertebrate food sources). These innovations could subsequently have been employed feeding on

the main food source (vegetation) in serving a useful scraping function for feeding on benthic algae from rocks. However, the oral and pharyngeal jaw teeth in *A. latilabris* also do not indicate an invertebrate diet, lacking the molariform teeth typical of mollusc-feeding thick-lipped species (e.g., Elmer *et al.* 2010). *Alcolapia latilabris* have long, slender oral teeth with broadened cusps and the crown bent at 90° inwards suggesting a specialised scraping function (Tichy & Seegers 1999). As such, it seems more likely that the thickened lips may be a primary adaptation to algal scraping, and that *A. latilabris* may be more ecologically similar to the specialised algal scrapers of the Great Lakes than invertebrate feeders with hypertrophied lips. In particular the specialised algae-scraping genera *Petrotilapia* from Lake Malawi and *Petrochromis* from Lake Tanganyika have broad, fleshy lips and have been suggested as examples of convergent evolution (Fryer & Iles 1972; Kassam *et al*. 2003). Although *Petrochromis* do not exhibit the hypertrophied lips or subterminal mouth of *A. latilabris*, they do have enlarged lips and shortened lower jaw with teeth visible even when the mouth is closed (Ribbink *et al*. 1983; Yamaota 1983; Lundeba *et al*. 2011). *Petrochromis macrognathus* in particular exhibits morphological similarity to *A. latilabris*, with retrognathous jaw, protruding upper jaw, concavity of the jaw region and pronounced and convexity of premaxillary ascending process (Yamaota 1983).

The oral teeth of *A. ndalalani* are markedly different from those of *A. latilabris*, being shorter and curved (rather than bent), perhaps indicative of a picking rather than scraping function, and were compared by (Tichy & Seegers 1999) to the oral dentition of short-snouted algae-eating Malawi cichlids *Labidochromis* (Lewis 1982). Thus, differences in lip and oral tooth morphology between *A. latilabris* and *A. ndalalani* (Tichy & Seegers 1999), but similar pharyngeal jaw morphology (Figure 5.28) and dietary signal (SIA, Figure 5.4), coupled with differences in particle size of stomach contents, suggest that these two species are feeding on and processing the same trophic resource (epilithic algae) but foraging in different ways, thus exploiting marginally differentiated trophic niches. Such fine-scale partitioning is not unusual in cichlid fishes, where co-occurring species can be differentiated based just on the cusping of oral teeth (Dieleman et al. 2015), and spatial segregation by foraging modes specialised to algae-scraping from either rocks or macrophytes (Bootsma *et al.* 1996).

As *A. latilabris* and *A. ndalalani* are resolved as sister species within the *Alcolapia* radiation (chapter three), occupy similar ecological niches (Figure 5.4) and currently occur in sympatry, the question arises as to whether their divergence could be a result of sympatric speciation. They certainly seem to meet at least three of the

four criteria proposed by (Coyne & Orr 2004; p. 142) for ascertaining sympatric speciation: i) species are sympatric; ii) substantial reproductive isolation, based on genetic differences, iii) taxa are sister groups, iv) existence of an allopatric phase unlikely based on biogeographic and evolutionary history. For the final criterion, it is not possible to conclusively rule out an allopatric history with secondary contact, given the fragmentary nature of the habitat and the extremely shallow lake allowing separation of pools and populations resulting from only small lake level changes. However, that the separation is recent enough to coincide with the geological separation of Lake Natron from Magadi (<10,000 years ago) suggests it is unlikely there was time for the species to emerge in isolation and subsequently repopulate the sites around the south of Lake Natron (although it has been suggested that cichlid fishes have previously diverged in as short a time as 200 years (Owen *et al.* 1990; Elmer *et al.* 2010). A similar, rapid, differentiation of two morphs ('incipient species') is observed in the Nicaraguan cichlids of Lake Apoyeque, where thin- and thick-lipped forms of *Amphilophus cf. citrinellus* arose in ~100 years (Elmer *et al.* 2010). The two forms also have significantly different body and head shape, with minimal genetic differentiation, but in contrast to the present study, the two morphs differ significantly in diet (stable isotope and gut content analysis).

Although not tested in the present study, the extent of reproductive isolation mediated by mate choice may also be an important factor in the divergence of sister species *A. latilabris* and *A. ndalalani*. All *Alcolapia* species exhibit sexual dichromatism, and male breeding colour is different between species. Assortative mating via female choice based on colour cues has been shown to be important in the divergence of incipient cichlid species (Seehausen *et al.* 2008; Selz *et al.* 2014). While allopatric sister species also frequently show colour differences driven by sexual selection, this is not typically accompanied by morphological differences (e.g., Kocher 2004; Salzburger 2009; Tyers & Turner 2013), and it is suggested that sexual selection coupled with natural selection may facilitate reproductive isolation in less ecologically differentiated morphs than divergent ecological selection alone (Seehausen & Wagner 2014).

Such apparently rapid changes in trophic morphology across the *Alcolapia* flock without corresponding genomic divergence (Figure 5.22) is indicative of adaptive radiation and filling of niche space within this system. Oral jaw morphology in cichlids is known to be largely controlled by only a few genes (Albertson *et al.* 2003), and adaptive divergence in jaw shape has been seen across the space of a few generations (van Rijssel *et al.* 2014). Furthermore, jaw and tooth shape are known to exhibit plastic responses to food type, and may thus quickly respond to habitat

changes. Finally, phenotypic plasticity can facilitate ecological diversification, by allowing novel phenotypes to survive in novel resource environments, and producing divergent selection pressures (Pfennig *et al.* 2010; Burress 2014).

Conclusions

An integrated dataset to investigate morphological and ecological differentiation between species provides further evidence of diversification within the *Alcolapia* group. The pattern of low genomic divergence with trophic and morphological differentiation is indicative of recent adaptive radiation with rapid ecological speciation. Future analyses may focus not only on the role of gene flow during the colonisation of and radiation within these lakes, but also the mechanics maintaining low genomic differentiation in concert with substantial ecomorphological divergence. There are several clear future goals to elucidate the pattern of ecomorphological differentiation within the system, including: determining regions of the genome responsible for morphological differentiation in these species; investigating the extent of reproductive isolation between Lake Natron species that occur in sympatry; testing the importance of sexual selection in this system based on male colouration; and determining the degree of trophic plasticity within *Alcolapia* based on resource use and substrate type.

References

Albertson RC, Kocher TD (2006) Genetic and developmental basis of cichlid trophic diversity. *Heredity*, **97**, 211–221.

Albertson RC, Streelman JT, Kocher TD (2003) Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 5252–5257.

Albrecht H, Apfelbach R, Wickler W (1968) Uber die Eigenstandigkeit der Art *Tilapia grahami* Boulenger, ihren Grubenbau und die Zucht in reines Susswasser (Pisces, Cichlidae). *Senckenberg Biologica*, **49**, 107–118.

Arbour JH, Lopez-Fernández H (2014) Adaptive landscape and functional diversity of Neotropical cichlids: implications for the ecology and evolution of Cichlinae (Cichlidae; Cichliformes). *Journal of Evolutionary Biology*, **27**, 2431–2442.

Arnold SJ (1983) Morphology, performance and fitness. *American Zoologist*, **23**, 347–361.

Astudillo-Clavijo V, Arbour JH, Lopez-Fernandez H (2015) Selection towards different adaptive optima drove the early diversification of locomotor phenotypes in the radiation of Neotropical geophagine cichlids. *BMC Evolutionary Biology*, **15**, 732.

Baumgarten L, Machado-Schiaffino G, Henning F, Meyer A (2015) What big lips are good for: On the adaptive function of repeatedly evolved hypertrophied lips of cichlid fishes. *Biological Journal of the Linnean Society*, **115**, 448–455.

Bootsma HA, Hecky RE, Hesslein RH, Turner GF (1996) Food partitioning among Lake Malawi nearshore fishes as revealed by stable isotope analyses. *Ecology*, **77**, 1286–1290.

Bravo GA, Remsen JV Jr., Brumfield RT (2014) Adaptive processes drive ecomorphological convergent evolution in antwrens (Thamnophilidae). *Evolution*, **68**, 2757–2774.

Burress ED (2014) Cichlid fishes as models of ecological diversification: patterns, mechanisms, and consequences. *Hydrobiologia*, **748**, 7-27

Coe MJ (1966) The Biology of *Tilapia grahami* Boulenger in Lake Magadi, Kenya. *Acta Tropica*, **23**, 146–198.

Colombo M, Diepeveen ET, Muschick M *et al.* (2012) The ecological and genetic basis of convergent thick-lipped phenotypes in cichlid fishes. *Molecular Ecology*, **22**, 670–684.

Correa C (2012) Tissue preservation biases in stable isotopes of fishes and molluscs from Patagonian lakes. *Journal of Fish Biology*, **81**, 2064–2073.

Correa C, Bravo AP, Hendry AP (2012) Reciprocal trophic niche shifts in native and invasive fish: salmonids and galaxiids in Patagonian lakes. *Freshwater Biology*, **57**, 1769–1781.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, MA.

Davis AM, Unmack PJ, Pusey BJ *et al.* (2013) Ontogenetic development of intestinal length and relationships to diet in an Australasian fish family (Terapontidae). *BMC Evolutionary Biology*, **13**, 1.

Davis AM, Unmack PJ, Pusey BJ *et al*. (2014) Evidence for a multi peak adaptive landscape in the evolution of trophic morphology in terapontid fishes. *Biological Journal of the Linnean Society*, **113**, 623–634.

DeNiro MJ, Epstein S (1977) Mechanism of carbon isotope fractionation associated with lipid synthesis. *Science*, **197**, 261–263.

Dieleman J, Van Bocxlaer B, Manntschke C *et al*. (2015) Tracing functional adaptation in African cichlid fishes through morphometric analysis of fossil teeth: exploring the methods. *Hydrobiologia*. doi:10.1007/s10750-015-2218-0.

Dray S, Dufour A-B (2007) The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**, 1–20.

Elmer KR, Lehtonen TK, Kautt AF, Harrod C, Meyer A (2010) Rapid sympatric ecological differentiation of crater lake cichlid fishes within historic times. *BMC Biology*, **8**, 60.

Firmat C, Schliewen UK, Losseau MEL, Alibert P (2012) Body shape differentiation at global and local geographic scales in the invasive cichlid *Oreochromis mossambicus. Biological Journal of the Linnean Society*, **105**, 369–381.

Ford AGP, Dasmahapatra KK, Rüber L *et al.* (2015) High levels of interspecific gene flow in an endemic cichlid fish adaptive radiation from an extreme lake environment. *Molecular Ecology*, doi: 10.1111–mec.13247.

France RL (1995) Carbon-13 enrichment in benthic compared to planktonic algae: foodweb implications. *Marine Ecology Progress Series*, **124**, 307–312.

Fryer G, Iles TD (1972) *The cichlid fishes of the Great Lakes of Africa: their biology and evolution.* Oliver & Boyd, Edinburgh.

Galis F, Drucker EG (1996) Pharyngeal biting mechanics in centrarchid and cichlid fishes: insights into a key evolutionary innovation. *Journal of Evolutionary Biology*, **9**, 641–670.

Gavrilets S, Losos JB (2009) Adaptive radiation: Contrasting theory with data. *Science*, **323**, 732–737.

Genner MJ, Turner GF, Barker S, Hawkins SJ (1999) Niche segregation among Lake Malawi cichlid fishes? Evidence from stable isotope signatures. *Ecology Letters*, **2**, 185–190.

Gomes JJ, Monteiro LR (2008) Morphological divergence patterns among populations of *Poecilia vivipara* (Teleostei Poeciliidae): test of an ecomorphological paradigm. *Biological Journal of the Linnean Society*, **93**, 799–812.

Goslee SC, Urban DL (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, **22**, 1–19.

Grant PR, Grant BR (2008) *How and why species multiply: the radiation of Darwin's finches*. Princeton Univ. Press, Princeton, NJ.

Grundler MC, Rabosky DL (2014) Trophic divergence despite morphological convergence in a continental radiation of snakes. *Proceedings of the Royal Society B*: *Biological Sciences*, **281**, 20140413–20140413.

Gunter HM, Fan S, Xiong F *et al.* (2013) Shaping development through mechanical strain: the transcriptional basis of diet-induced phenotypic plasticity in a cichlid fish. *Molecular Ecology*, **22**, 4516–4531.

Hamilton SK, Lewis WMJ (1992) Stable carbon and nitrogen isotopes in algae and detritus from the Orinco River floodplain, Venezuela. *Geochimica et Cosmochimica Acta*, **56**, 4237–4246.

Hammer O, Harper DAT, Ryan PD (2001) PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, **4**, 1–9.

Hjelm J, Svanback R, Bystrom P *et al*. (2001) Diet-dependent body morphology and ontogenetic reaction norms in Eurasian perch. *Oikos*, **95**, 311–323.

Hulsey CD (2006) Function of a key morphological innovation: fusion of the cichlid pharyngeal jaw. *Proceedings of the Royal Society B*, **273**, 669–675.

Hulsey CD, Roberts RJ, Lin ASP *et al*. (2008) Convergence in a mechanically complex phenotype: Detecting structural adaptations for crushing in cichlid fish. *Evolution*, **62**, 1587–1599.

Ingram T, Mahler DL (2013) SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods in Ecology and Evolution*, **4**, 416–425.

Irschick DJ, Meyers JJ, Husak JF, Le Galliard J-F (2008) How does selection operate on whole-organism functional performance capacities? A review and synthesis. *Evolutionary Ecology Research*, **10**, 177–196.

Jackson AL, Inger R, Parnell AC, Bearhop S (2011) Comparing isotopic niche widths among and within communities: SIBER - Stable Isotope Bayesian Ellipses in R. *The Journal of animal ecology*, **80**, 595–602.

Jackson MC, Donohue I, Jackson AL *et al.* (2012) Population-level metrics of trophic structure based on stable isotopes and their application to invasion ecology. *PLoS ONE*, **7**, e31757.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Kaliontzopoulou A, Carretero MA, Llorente GA (2010) Intraspecific ecomorphological variation: linear and geometric morphometrics reveal habitat-related patterns within *Podarcis bocagei* wall lizards. *Journal of Evolutionary Biology*, **23**, 1234–1244.

Kassam DD, Adams DC, Hori M, Yamaoka K (2003) Morphometric analysis on ecomorphologically equivalent cichlid species from Lakes Malawi and Tanganyika. *Journal of Zoology*, **260**, 153–157.

Kelly B, Dempson JB, Power M (2006) The effects of preservation on fish tissue stable isotope signatures. *Journal of Fish Biology*, **69**, 1595–1611.

Kiljunen M, Grey J, Sinisalo T *et al.* (2006) A revised model for lipid-normalizing $\delta^{13}C$ values from aquatic organisms, with implications for isotope mixing models. *Journal of Applied Ecology*, **43**, 1213–1222.

Klingenberg CP (2011) MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources*, **11**, 353–357.

Klingenberg CP, Gidaszewski NA (2010) Testing and quantifying phylogenetic signals and homoplasy in morphometric data. *Systematic Biology*, **59**, 245–261.

Klingenberg CP, Barluenga M, Meyer A (2003) Body shape variation in cichlid fishes of the *Amphilophus citrinellus* species complex. *Biological Journal of the Linnean Society*, **80**, 397–408.

Klingenberg CP, Duttke S, Whelan S, Kim M (2011) Developmental plasticity, morphological variation and evolvability: a multilevel analysis of morphometric integration in the shape of compound leaves. *Journal of Evolutionary Biology*, **25**, 115–129.

Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. *Nature Reviews Genetics*, **5**, 288–298.

Kramer DL, Bryant MJ (1995) Intestine length in the fishes of a tropical stream: Relationships to diet - the long and short of a convoluted issue. *Environmental Biology of Fishes*, **42**, 129–141.

Layman CA, Arrington DA, Montan CG, Post DG (2007) Can stable isotope ratios provide for community-wide measures of trophic structure? *Ecology*, **88**, 42–48.

Lewis DC (1982) A revision of the genus *Labidochromis* (Teleostei: Cichlidae) from Lake Malawi. *Zoological Journal of the Linnean Society*, **75**, 189–265.

Liem KF (1973) Evolutionary strategies and morphological innovations: Cichlid pharyngeal jaws. *Systematic zoology*, **22**, 425.

Losos JB, Irschick DJ, Schoener TW (1994) Adaptation and constraint in the evolution of specialization of Bahamian *Anolis* lizards. *Evolution*, **48**, 1786–1798.

Lundeba M, Stauffer JR Jr, Konings AF (2011). Five new species of the genus *Petrotilapia* (Teleostei: Cichlidae), from Lake Malawi, Africa. *Ichthyological Explorations of Freshwaters*, **22**, 149–168.

Maddison WP, Maddison DR (2014) Mesquite: a modular system for evolutionary analysis. Version 3.02. *mesquiteproject.org*.

Magalhaes IS, Lundsgaard-Hansen B, Mwaiko S, Seehausen O (2012) Evolutionary divergence in replicate pairs of ecotypes of Lake Victoria cichlid fish. *Evolutionary Ecology Research*, **14**, 381–401.

Mahler DL, Ingram T, Revell LJ, Losos JB (2013) Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science*, **341**, 290–292.

Malhotra A, Thorpe RS (1991) Microgeographic variation in *Anolis oculatus*, on the island of Dominica, West Indies. *Journal of Evolutionary Biology*, **4**, 321–335.

Manousaki T, Hull PM, Kusche H *et al.* (2013) Parsing parallel evolution: ecological divergence and differential gene expression in the adaptive radiations of thick-lipped Midas cichlid fishes from Nicaragua. *Molecular Ecology*, **22**, 650–669.

Marijnissen SAE, Michel E, Cleary DFR, McIntyre PB (2008) Ecology and conservation status of endemic freshwater crabs in Lake Tanganyika, Africa. *Biodiversity and Conservation*, **18**, 1555–1573.

Martin CH (2013) Strong assortative mating by diet, color, size, and morphology but limited progress toward sympatric speciation in a classic example: Cameroon crater lake cichlids. *Evolution*, **67**, 2114–2123.

Moeri O, da Silveira Lobo Sternberg L, Rodicio LP, Walsh PJ (2003) Direct effects of ambient ammonia on the nitrogen isotope ratios of fish tissues. *Journal of Experimental Marine Biology and Ecology*, **282**, 61–66.

Muschick M, Barluenga M, Salzburger W, Meyer A (2011) Adaptive phenotypic plasticity in the Midas cichlid fish pharyngeal jaw and its relevance in adaptive radiation. *BMC Evolutionary Biology*, **11**, 116.

Muschick M, Indermaur A, Salzburger W (2012) Convergent evolution within an adaptive radiation of cichlid fishes. *Current Biology*, **22**, 1–7.

Muschick M, Nosil P, Roesti M *et al.* (2014) Testing the stages model in the adaptive radiation of cichlid fishes in East African Lake Tanganyika. *Proceedings of the Royal Society B: Biological Sciences*, **281**, 20140605.

Ogle DH (2015) FSA: Fisheries Stock Analysis. *R package version 0.6.16.*

Owen RB, Crossley R, Johnson TC *et al.* (1990) Major low lake levels of Lake Malawi and their implications for speciation rates in cichlid fishes. *Proceedings of the Royal Society B: Biological Sciences*, **240**, 519–553.

Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

Parnell AC, Inger R, Bearhop S, Jackson AL (2010) Source partitioning using stable isotopes: coping with too much variation. *PLoS ONE*, **5**, e9672.

Perkins MJ, McDonald RA, van Veen FJF *et al.* (2014) Application of nitrogen and narbon stable isotopes (δ15N and δ13C) to quantify food chain length and trophic structure. *PLoS ONE*, **9**, e93281.

Pfennig DW, Wund MA, Snell-Rood EC *et al.* (2010) Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology & Evolution*, **25**, 459–467.

Pinnegar JK, Polunin NVC (1999) Differential fractionation of $\delta^{13}C$ and $\delta^{15}N$ among fish tissues: implication for the study of trophic interactions. *Functional Ecology*, **13**, 225–231.

Polly PD (2014) Geometric morphometrics for Mathematica. Version 11.0.

Polly PD, Lawing AM, Fabre A-C, Goswami A (2013) Phylogenetic principal components analysis and geometric morphometrics. *Hystrix*, **24**, 33–41.

Post DM (2002) Using stable isotopes to estimate trophic position: Models, methods, and assumptions. *Ecology*, **83**, 703–718.

Post DM (2003) Individual variation in the timing of ontogenetic shifts in largemouth bass. *Ecology*, **84**, 1298–1310.

Praebel K, Knudsen R, Siwertsson A *et al.* (2013) Ecological speciation in postglacial European whitefish: Rapid adaptive radiations into the littoral, pelagic, and profundal lake habitats. *Ecology and Evolution*, **3**, 4970–4986.

Ribbink AJ, Marsh BA, Marsh AC, Ribbink AC, Sharp BJ (1983) A preliminary survey of the cihlid fishes of rocky shore habitats in Lake Malawi. *South African Journal of Zoology*, **18**, 149–310.

Rohlf FJ (2013) tpsDig version 2.17. Department of Ecology and Evolution, State University of New York at Stony Brook, New York.

Rohlf FJ, Slice D (1990) Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic zoology*, **39**, 40–59.

Rüber L, Adams DC (2001) Evolutionary convergence of body shape and trophic morphology in cichlids from Lake Tanganyika. *Journal of Evolutionary Biology*, **14**, 325–332.

Rüber L, Verheyen E, Meyer A (1999) Replicated evolution of trophic specializations in an endemic cichlid fish lineage from Lake Tanganyika. *Proceedings of the National Academy of Sciences of the United States of America*, **96**, 10230–10235.

Saitou N, Nei M (1987) Reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.

Salzburger W (2009) The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Molecular Ecology*, **18**, 169–185.

Schliewen UK, Klee B (2004) Reticulate sympatric speciation in Cameroonian crater lake cichlids. *Frontiers in Zoology*, **1**, 5.

Schluter D (1994) Experimental evidence that competition promotes divergence in adaptive radiation. *Science*, **266**, 798–801.

Schluter D (2000) *The ecology of adaptive radiation*. Oxford University Press, Oxford.

Schluter D, McPhail JD (1992) Ecological character displacement and speciation in sticklebacks. *The American Naturalist*, **140**, 85–108.

Schoener TW (1970) Nonsynchronous spatial overlap of lizards in patchy habitats. *Ecology*, **51**, 408–418.

Seegers L, Tichy H (1999) The *Oreochromis alcalicus* flock (Teleostei: Cichlidae) from Lake Natron and Magadi, Tanzania and Kenya, with description of two new species. *Ichthyological Explorations of Freshwaters*, **10**, 97–146.

Seehausen O, Wagner CE (2014) Speciation in freshwater fishes. *Annual Review of Ecology, Evolution, and Systematics*, **45**, 621–651.

Seehausen O, Terai Y, Magalhaes IS *et al.* (2008) Speciation through sensory drive in cichlid fish. *Nature*, **455**, 620–626.

Selz OM, Pierotti MER, Maan ME, Schmid C, Seehausen O (2014) Female preference for male color is necessary and sufficient for assortative mating in 2 cichlid sister species. *Behavioral Ecology*, **25**, 612–626.

Snorrason SS, Skulason S (2004) Adaptive speciation in northern freshwater fish—patterns and processes. In: *Adaptive Speciation* (eds Dieckmann U, Doebeli M, Metz JAJ, Tautz D). Cambridge University Press, Cambridge.

Snowberg LK, Bolnick DI (2012) Partitioning the effects of spatial isolation, nest habitat, and individual diet in causing assortative mating within a population of threespine stickleback. *Evolution*, **66**, 3582–3594.

Solomon CT, Carpenter SR, Clayton MK *et al.* (2011) Terrestrial, benthic, and pelagic resource use in lakes: results from a three-isotope Bayesian mixing model. *Ecology*, **92**, 1115–1125.

Sweeting CJ, Polunin NVC, Jennings S (2006) Effects of chemical lipid extraction and arithmetic lipid correction on stable isotope ratios of fish tissues. *Rapid Communications in Mass Spectrometry*, **20**, 595–601.

Theis A, Ronco F, Indermaur A, Salzburger W, Egger B (2014) Adaptive divergence between lake and stream populations of an East African cichlid fish. *Molecular Ecology*, **23**, 5304–5322.

Tichy H, Seegers L (1999) The *Oreochromis alcalicus* flock (Teleostei Cichlidae) from lakes Natron and Magadi, Tanzania and Kenya: A model for the evolution of new species flocks in historical times. *Ichthyological Explorations of Freshwaters*, **10**, 147–174.

Trewavas E (1983) *Tilapiine Fishes of the genera Sarotherodon, Oreochromis and Danakilia*. British Museum (Natural History), London.

Tyers AM, Turner GF (2013) Signal and preference divergence among populations of the non-endemic basal Lake Malawi cichlid fish *Astatotilapia calliptera* (Perciformes: Cichlidae). *Biological Journal of the Linnean Society*, **110**, 180–188.

van Rijssel JC, Hoogwater ES, Kishe-Machumu MA *et al.* (2014) Fast adaptive responses in the oral jaw of Lake Victoria cichlids. *Evolution*, **69**, 179–189.

Vermeij GJ (1973) Adaptation, versatility, and evolution. *Systematic zoology*, **22**, 466–477.

Vizza C, Sanderson BL, Burrows DG, Coe HJ (2013) The effects of ethanol preservation on fish fin stable isotopes: Does variation in C:N ratio and body size matter? *Transactions of the American Fisheries Society*, **142**, 1469–1476.

Wagner CE, McIntyre PB, Buels KS *et al*. (2009) Diet predicts intestine length in Lake Tanganyika's cichlid fishes. *Functional Ecology*, **23**, 1122–1131.

Wallace RK (1981) An assessment of diet-overlap indexes. *Transactions of the American Fisheries Society*, **110**, 72–76.

Ward JHJ (1963) Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, **58**, 236–244.

Wilson PJ, Wood CM, Walsh PJ *et al.* (2004) Discordance between genetic structure and morphological, ecological, and physiological adaptation in Lake Magadi tilapia. *Physiological and Biochemical Zoology*, **77**, 537–555.

Yamaoka K (1983) A revision of the cichlid fish genus *Petrochromis* from Lake Tanganyika, with description of a new species. *Japanese Journal of Ichthyology*, **30**, 129–141.

Zaccara S, Crosa G, Vanetti I *et al.* (2014) Genetic and morphological analyses indicate high population mixing in the endangered cichlid *Alcolapia* flock of East Africa. *Conservation Genetics*, **15**, 429–440.

Appendix 5

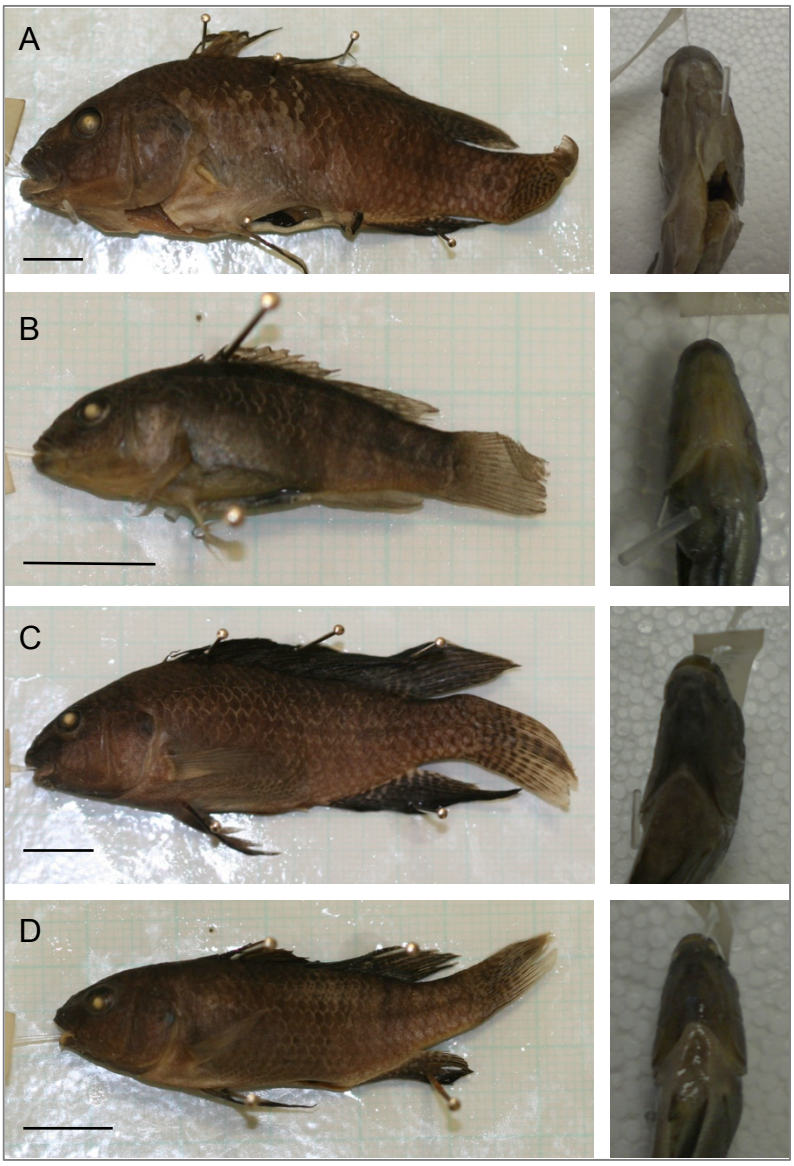


Figure 5A.1. Photographs of site 17 *A. alcalica* specimens.

Lateral (left) and ventral (right-hand column) photographs of specimens initially identified as *A. alcalica* from site 17. Specimens shown in panels C and D were later amended to *A.* aff. *ndalalani* on re-examination based on the shorter lower jaw relative to upper jaw. Scale bar: 10mm.

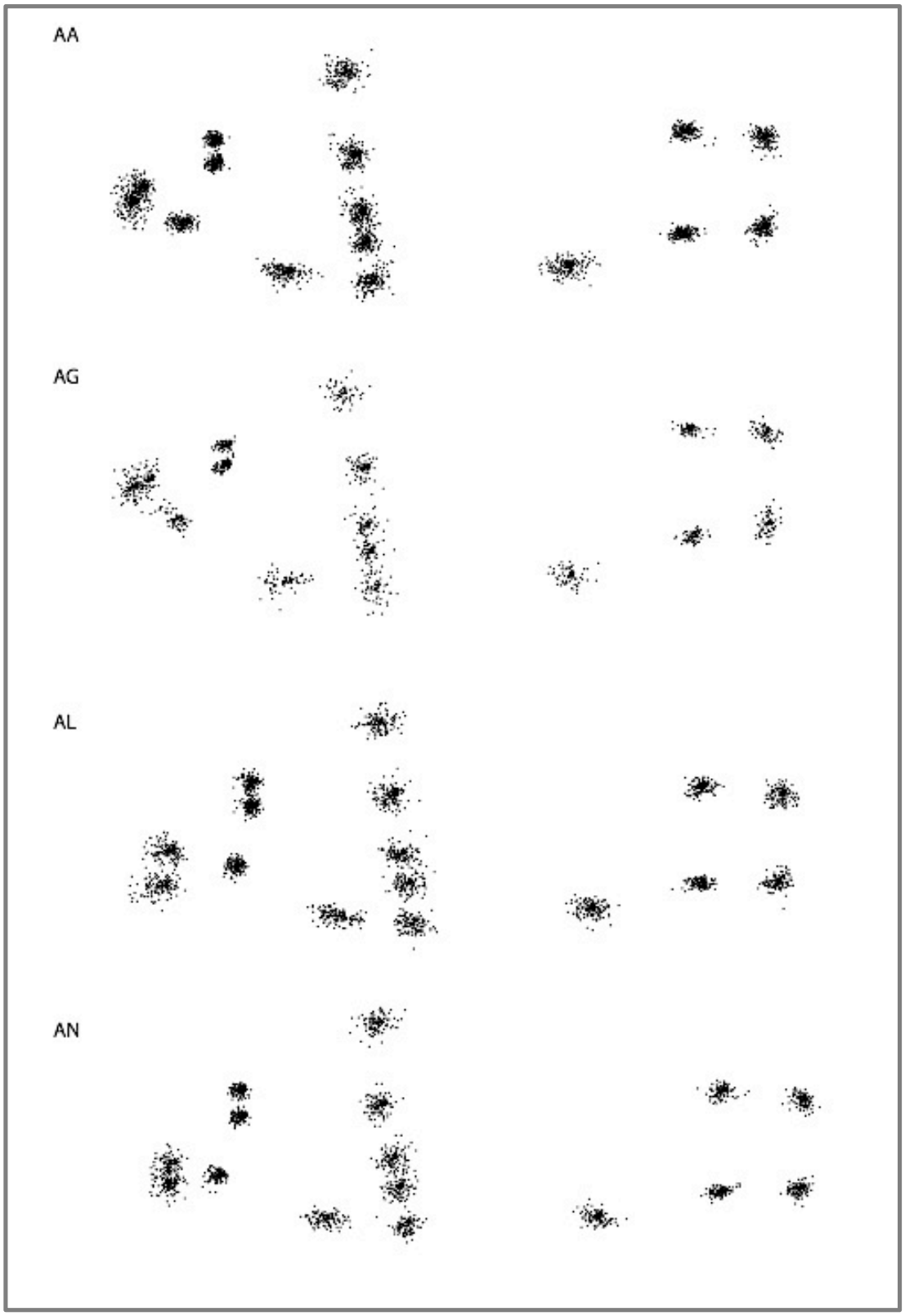


Figure 5A.2. Variation of landmark data for RAD single-species datasets only.

AA: *A. alcalica*; AG: *A. grahami*; AL: *A. latilabris*; AN: *A. ndalalani*. Scatter around the mouth landmarks appears unimodal for each species (rather than bimodal as observed in the combined dataset for all species).

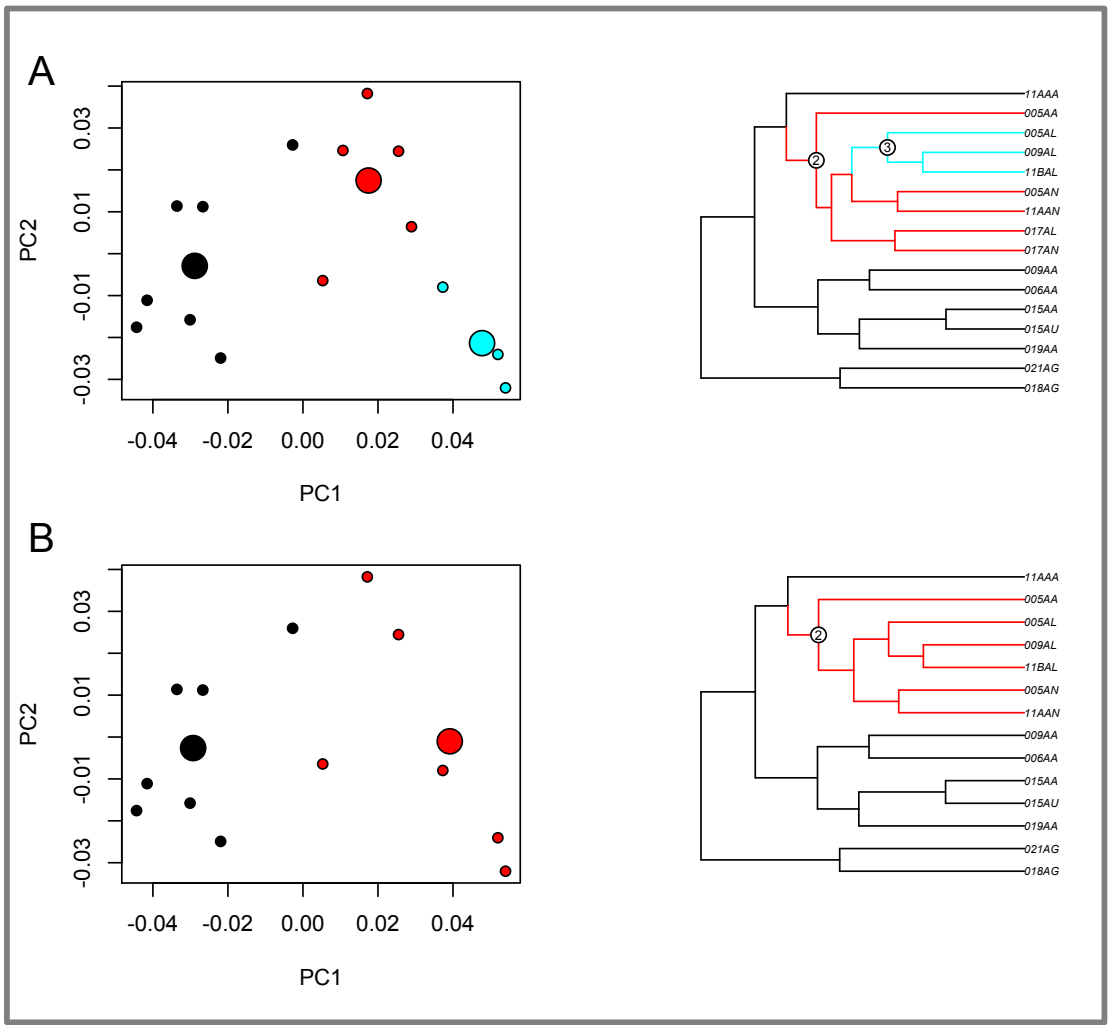


Figure 5A.3. SURFACE analysis for reduced datasets.

A) excluding *A. alcalica* population from site 17; B) excluding all populations from site 17.

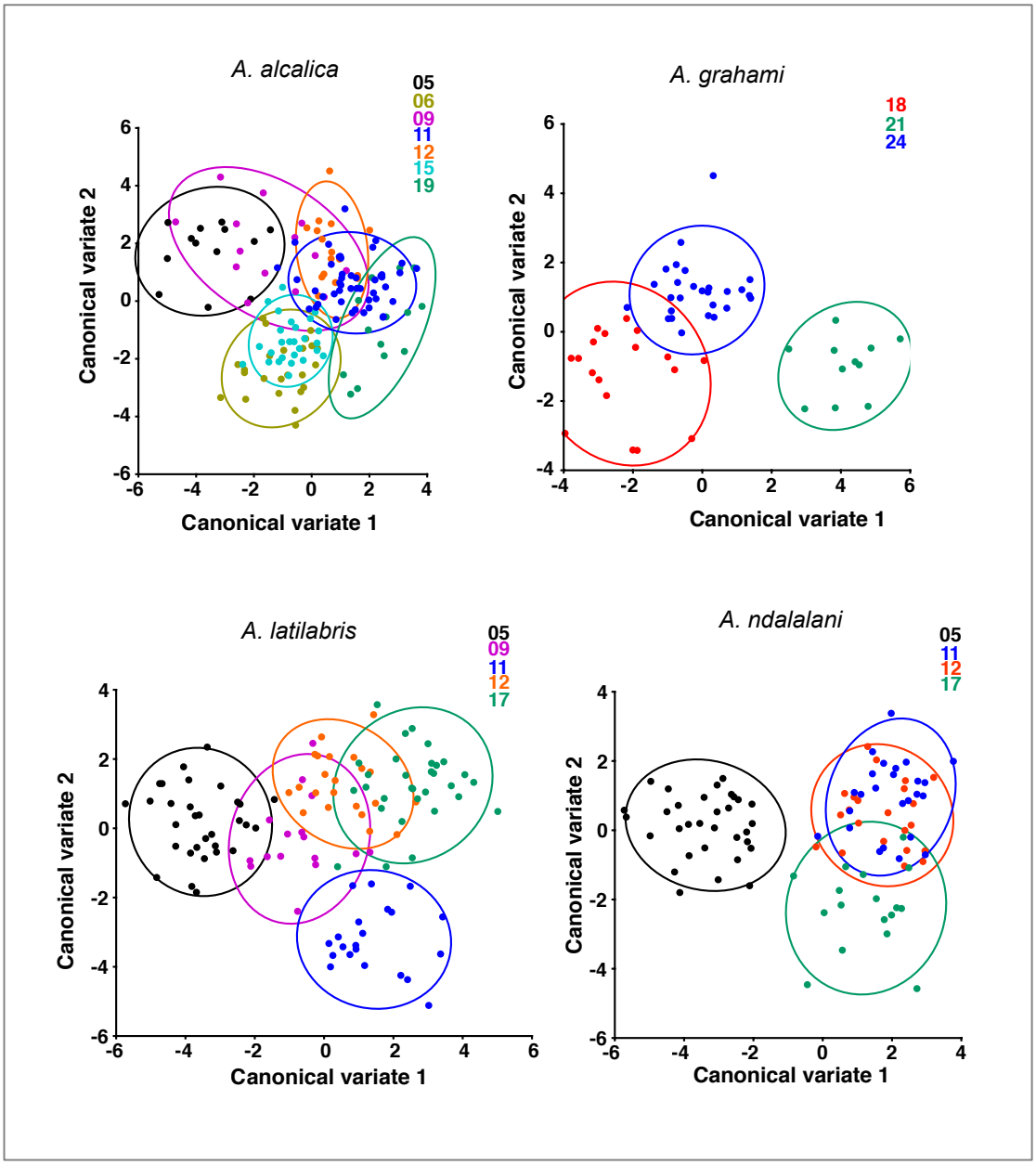


Figure 5A.4. **Canonical variate analysis plots for single-species datasets** (used to generate morphological distances for Mantel tests). Legend indicates sampling site (population). Full description of intraspecific variation and associated shape changes is included in Chapter Six.

Table 5A.1. Eigenvalue results from PCA across the entire body shape dataset. Values across A) all samples; and B) population means for site/species groups.

PC	A			B		
	Eigenvalue	Var (%)	Cumul (%)	Eigenvalue	Var (%)	Cumul (%)
1	0.001366	38.14	38.14	0.000997	55.46	55.46
2	0.000817	22.81	60.96	0.000454	25.28	80.75
3	0.000201	5.62	66.57	0.000102	5.67	86.41
4	0.000184	5.13	71.70	0.000069	3.83	90.24
5	0.000144	4.03	75.73	0.000038	2.11	92.35
6	0.000135	3.77	79.50	0.000033	1.85	94.20
7	0.000117	3.27	82.77	0.000030	1.65	95.86
8	0.000112	3.13	85.89	0.000023	1.27	97.12
9	0.000083	2.31	88.21	0.000017	0.93	98.06
10	0.000069	1.92	90.12	0.000008	0.44	98.50
11	0.000055	1.55	91.67	0.000006	0.33	98.83
x	0.000045	1.25	92.91	0.000006	0.31	99.13
13	0.000039	1.09	94.00	0.000004	0.23	99.37
14	0.000035	0.98	94.98	0.000003	0.18	99.55
15	0.000031	0.86	95.85	0.000003	0.15	99.70
16	0.000027	0.76	96.61	0.000002	0.11	99.81
17	0.000021	0.58	97.18	0.000001	0.06	99.87
18	0.000016	0.46	97.64	0.000001	0.05	99.92
19	0.000014	0.38	98.02	0.000001	0.03	99.95
20	0.000012	0.34	98.36	0.000000	0.02	99.97
21	0.000012	0.33	98.69	0.000000	0.02	99.99
22	0.000011	0.30	98.99	0.000000	0.01	99.99
23	0.000009	0.25	99.24	0.000000	0.00	100.00
24	0.000008	0.23	99.47	0.000000	0.00	100.00
25	0.000007	0.20	99.67	-	-	-
26	0.000006	0.16	99.83	-	-	-
27	0.000005	0.14	99.97	-	-	-
28	0.000001	0.03	100.00	-	-	-

Cumul: Cumulative variance; Var: variance.

Table 5A.2. Eigenvalue results from PCA across the PHJ dataset.

PC	Eigenvalues	% Variance	% Cumulative
1	**0.00209386**	**47.87**	**47.87**
2	**0.00067336**	**15.40**	**63.27**
3	**0.00060854**	**13.91**	**77.18**
4	**0.00037546**	**8.58**	**85.77**
5	**0.00023366**	**5.34**	**91.11**
6	**0.00013502**	**3.09**	**94.19**
7	**0.00010345**	**2.37**	**96.56**
8	**0.00007477**	**1.71**	**98.27**
9	**0.00004295**	**0.98**	**99.25**
10	**0.00003275**	**0.75**	**100.00**

Table 5A.3. Eigenvalue results from PCA across the PHJ dataset at sites 05 and 12 only.

	Eigenvalues	% Variance	Cumulative %
1	0.00267015	51.62	51.62
2	0.0008757	16.93	68.55
3	0.00058469	11.30	79.86
4	0.0004445	8.59	88.45
5	0.00021019	4.06	92.51
6	0.00013873	2.68	95.19
7	0.00010405	2.01	97.21
8	0.00006385	1.23	98.44
9	0.00004786	0.93	99.37
10	0.00003284	0.64	100.00

Table 5A.4. Results of simple and partial Mantel tests (R-values) for isolation by adaptation.

Test	Simple Mantel (Morph ~ Geography)				Simple Mantel (Morph ~ RAD)		Partial Mantel (Morph ~ RAD + Geography)			
Geographic distance	Lake Perimeter		Straight-line				Lake Perimeter		Straight-line	
Morphometric distance	Procrustes	Mahalanobis	Procrustes	Mahalanobis	Procrustes	Mahalanobis	Procrustes	Mahalanobis	Procrustes	Mahalanobis
A. alcalica	0.0199	0.0477	0.0515	0.1448	0.1258	0.1675	0.1277	0.1719	0.1349	0.1925
A. grahami	-	-	0.8802	-0.9128	0.1273	-0.8638	-	-	-1.00	-1.00
A. latilabris	0.1921	-0.0987	0.1921	-0.0987	0.1046	-0.2132	0.1046	-0.1977	0.1284	0.1284
A. ndalalani	0.2682	0.272	0.2156	0.2402	0.7523	0.8062*	0.8500*	0.9293*	0.8929*	0.9524*

*indicates significant at P<0.05)

Chapter six

Intraspecific variation of soda lake cichlid ecomorphology

Abstract

Intraspecific analysis provides information on how species react to divergent selection pressures in different environmental conditions. Chapter six examines the ecological and morphological divergence of the *Alcolapia* species flock at an intraspecific population level to consider ecomorphological differentiation over geographic scales. Niche breadth (based on stable isotopes and stomach contents analysis) differs between populations of *A. alcalica* occurring allopatrically or sympatrically with other species, suggesting a role for competition in niche exploitation. Conversely, niche breadth is constrained in *A. grahami* across its range (including translocated populations), suggesting niche conservatism in different environmental conditions. Populations for all species are mostly strongly differentiated in body shape, with the exception of populations of *A. ndalalani*, suggesting reduced variability of this species across its (limited) geographic range. The inclusion of additional populations with ecomorphological data allowed identification of another possible population of hybrid origin with individuals displaying intermediate ecomorphological values for *A. latilabris* and *A. alcalica* (see also discussion in previous chapters on hybridisation at site 17). The analysis of sex ratios in several populations also identifies a possible case of sexual dimorphism of an ecological trait (trophic morphology) from one of the *A. alcalica* populations. Finally, while intraspecific body size varied by population, it was not correlated with environmental measures. These results suggest that the geographical separation of *Alcolapia* populations across their distribution results in substantial ecomorphological divergence, and raises the possibility that these populations are undergoing adaptive divergence despite shallow genomic divergence and ongoing gene flow (chapter three).

Introduction

Studying species across their geographic range provides information on species adaptation to varying environmental conditions. Such variation of environmental conditions between habitats constitutes a source of divergent natural selection between populations, as individuals will experience different adaptive landscapes between environments (Nosil 2012). Adaptive divergence to different environments may be a precursor to speciation via ecological or mutation-order mechanisms (Schluter 2009). Considering cases of such local adaptation may thus inform on processes promoting or constraining speciation across spatial scales. Local adaptation and adaptive divergence across species ranges is seen in many systems including body shape morphology and life history between lake / stream morphs of postglacial fishes (Schluter 1993; Østbye *et al.* 2005; Moser *et al.* 2012); morphology and foraging mode of Galapagos finches (Kleindorfer *et al.* 2006); and vocalisation divergence along elevational gradients in tropical passerines (Caro *et al.* 2013). Such divergence is also observed over microgeographic scales, with subdivisions within populations, including sparid fish populations between lagoons in a single estuary (Cutwa & Turingan 2000), depth gradients in lacustrine cichlids (Barluenga *et al.* 2006; Seehausen *et al.* 2008), and between host tree stands in island scrub jays (Langin *et al.* 2015).

Gene flow can constrain the extent of local adaptation, which may make phenotypic plasticity important in local divergence. The interplay of gene flow between populations vs. the strength of diversifying selection between environments may dictate how far along the speciation continuum adaptive divergence proceeds. Gene flow may prevent local adaptation through homogenising effects and prevent the development of reproductive isolation, but conversely may also contribute to adaptive divergence by facilitating shifts between adaptive peaks (Hendry *et al.* 2002 and refs therein). The maintenance of locally adapted phenotypes despite extensive gene flow from divergent populations has been demonstrated in Trinidadian guppies (Fitzpatrick *et al.* 2015). Furthermore, it has been proposed that intraspecific variation can generate novel phenotypes through hybridisation on secondary contact (Nichols *et al.* 2014). The Natron-Magadi *Alcolapia* populations are separated by trona crust but also experience moderate levels of migration between populations, possibly during flooding, and experience ongoing gene flow (Zaccara *et al.* 2014; Ford *et al.* 2015, see also chapter three). These conditions predispose populations to local adaptation via divergent selection on environmental gradients in allopatrically separated populations, but also provide increased genetic

variation on which natural selection can act via infrequent migration between populations.

Interspecific morphological variation may exhibit a mismatch of phenotype to environment due to a correlation between the lineage's ancestral habitat and evolved morphology. Such ancestral morphologies may be sustained after environmental changes due to genetic or phylogenetic constraints (Liem 1980; Binning & Chapman 2010). As such, considering intraspecific differences may better allow ecological selection to be differentiated without the influence of phylogenetic constraint (Futuyma & Moreno 1988; Binning & Chapman 2010). This may be particularly pertinent to the soda lake system given that geological evidence suggests a considerably different environment in the recent past (10 KYA; Roberts *et al.* 1993), being deeper and less saline-alkaline.

The soda lakes Natron and Magadi present an intriguing setting in which to investigate intraspecific variation. Although there are not obvious paired populations between which ecological divergence is seen in other fishes (such as benthic/pelagic, river/lake freshwater/marine; e.g., Schluter 1993; Jones *et al.* 2012; Hulsey *et al.* 2013), the springs and lagoons differ substantially in hydrochemistry, flow and species abundance/occurrence (pers. obs.). Thus, rather than offering a dichotomous environmental comparison, the Natron-Magadi populations offer an environmental gradient across populations in which to investigate ecomorphological differentiation. Environmental conditions tested in this study are given in Appendix Table 6A.1. Furthermore, *A. grahami* in Lake Magadi exhibit local morphological, physiological and behavioural differences between populations (Wilson et al. 2004), suggesting localised adaptation. The strength of divergent selection between environmental regimes is typically measured based on individual fitness, and examined using reciprocal translocation experiments between environments (Nosil 2012). While individual fitness is not explicitly tested here, the translocation of two populations of *A. grahami* to Kenyan lakes (Nakuru and Elementeita) in the 1950s (Hickley *et al.* 2008) provides a natural experiment, allowing short-term (~100s generations) local adaptation to a novel environment to be examined.

Sexual dichromatism and sexual size dimorphism have commonly been reported in cichlids (e.g., Erlandsson & Ribbink 1997; Schütz & Taborsky 2000; Maan & Sefc 2013), however dimorphism of shape has been less well documented. Trewavas (1983) notes sexual dimorphism in *Oreochromis* and *Alcolapia* species, however, this mainly relates to soft-tissue or internal features: genital papilla; male elongation of soft dorsal and anal fins, and simplification of oral teeth in males. In some *Oreochromis* species, there is an elongation of the jaws in male specimens

(Trewavas 1983), although females of *O. mossambicus* have been reported to exhibit wider premaxilla and longer snouts than males of the same species (Oliveira & Almada 1995). *Tropheus* populations from Lake Tanganyika (*T. polli* and *T. moori*) exhibit sexual dimorphism in cranial shape, with females exhibiting a larger buccal area (attributed as an adaptation to mouthbrooding), while different populations of the same species exhibited differences of mouth position unrelated to sex, which were attributed to different ecological selection regimes between habitats (Herler *et al.* 2010). However, sexual dimorphism of mouth position has been reported in populations of *Astatotilapia burtoni* from Lake Tanganyika (Theis *et al.* 2014). While sexual dichromatism of *Alcolapia* was discussed extensively in the most recent taxonomic revision (Seegers & Tichy 1999), to my knowledge there are no existing accounts of sexual dimorphism (except total size) within the species. Skewed sex ratios have previously been reported in *A. grahami*, which was thought to be a result of the environmental conditions at Lake Magadi (Papah *et al.* 2013). The same study demonstrated that *A. grahami* spermatozoa exhibit adaptations to the extreme conditions, so appear to be tolerant of the pH and salinity experienced in external fertilisation (Papah *et al.* 2013). Closely related *Oreochromis niloticus* exhibits sex determination dependent on genetic factors and environmental factors (temperature but not salinity) during development (Abucay *et al.* 1999; Bezault *et al.* 2007). However, sex ratios are also male-biased in other *Oreochromis* species occurring at moderate temperatures in freshwater (e.g., *O. variabilis;* Maithya *et al.* 2012). To investigate patterns of sexual dimorphism and skewed sex ratios within the *Alcolapia* species, this chapter also considers the effects of sex alongside the ecological and morphological characterisation of intraspecific variation.

Aims

Here, intraspecific ecological and morphological differentiation is investigated within *Alcolapia* species to test the extent of adaptive divergence among populations of soda lake cichlids. Ecomorphological data are included from an additional six sampling sites to those in chapter five (four in Lake Natron, one in Lake Magadi, and one in Lake Elementeita), for which genomic (RAD) data were not available. Intraspecific relationships within each species are explored to consider local adaptation and the correlation of phenotype with environment.

The soda lakes offer an intriguing setting in which to study phylogeographic structure and local adaptation of recently diverged species. The presence of several

species with differing distributions allow several hypotheses to be tested: i) Do *Alcolapia* species in Natron show a broader niche range than Magadi? This may be predicted due to increased total area and geographical range of species, increased allochthonous input, larger variation in resource availability, increased ecological niche space, and several inflowing rivers/streams; ii) Does the ubiquitous *A. alcalica* show a greater extent of variability than species constrained to southern Natron? iii) Do species in sympatry exhibit a narrower range of niche exploitation than where they occur in single-species groups? iv) Does ecological niche exploitation exhibit plasticity when species are translocated to new environments (Lake Nakuru and Elementeita), or is the ecological niche conserved? v) Is local adaptation (and adaptive divergence) observed in the face of high levels of gene flow?

Methods

Sampling

Sampling and tissue preservation were conducted as described in chapter five methods. Total specimen numbers by analysis for the current study are given in Table 6.1, along with site information and GPS co-ordinates. Sampling locations are displayed on the map in Figure 6.1.

Table 6.1. Sampling locations and specimen numbers by analysis.

GMM: Geometric morphometrics; SIA: Stable isotope analysis.

Lake	Site	Site GPS Coordinates	Species	SIA	Stomach contents	GMM: Body
Natron	5	-2.5976, 35.9184	*A. alcalica*	13	10	13
			A. latilabris	15	13	31
			A. ndalalani	12	11	32
	6	-2.4304, 35.8954	*A. alcalica*	15	-	24
	7	-2.4413, 35.8934	*A. alcalica*	20	-	12
	8	-2.4713, 35.8879	*A. alcalica*	13	-	16
	9	-2.4713, 35.8879	*A. alcalica*	16	-	14
			A. latilabris	15	-	16
	11	-2.5910, 36.0010	*A. alcalica*	29	10	73
			A. latilabris	19	12	66
			A. ndalalani	30	12	63
	12	-2.6190, 35.9998	*A. alcalica*	15	12	16
			A. latilabris	15	10	22
			A. ndalalani	15	11	22
	13	-2.5994, 35.9112	*A. alcalica*	15	-	19
			A. latilabris	15	-	18
			A. ndalalani	-	-	4
	14	-2.5270, 36.0461	*A. alcalica*	15	-	33
			A. latilabris	17	-	33
	15	-2.4334, 36.1018	*A. alcalica*	15	12	27
			A. alcalica (upturned)	16	28	30
	17	-2.4563, 36.0878	*A. alcalica*	4	-	-
			A. latilabris	16	-	31
			A. ndalalani	14	-	17
	19	-2.1458, 36.0558	*A. alcalica*	13	12	14
Magadi	18	-2.0011, 36.2320	*A. grahami*	16	11	18
	21	-1.8444, 36.2243	*A. grahami*	15	10	11
	22	-1.8927, 36.2648	*A. grahami*	15	-	15
Elementeita	23	-0.4791, 36.2558	*A. grahami*	15	-	17
Nakuru	24	-0.3960, 36.1076	*A. grahami*	15	11	27
			Total	**458**	**185**	**734**

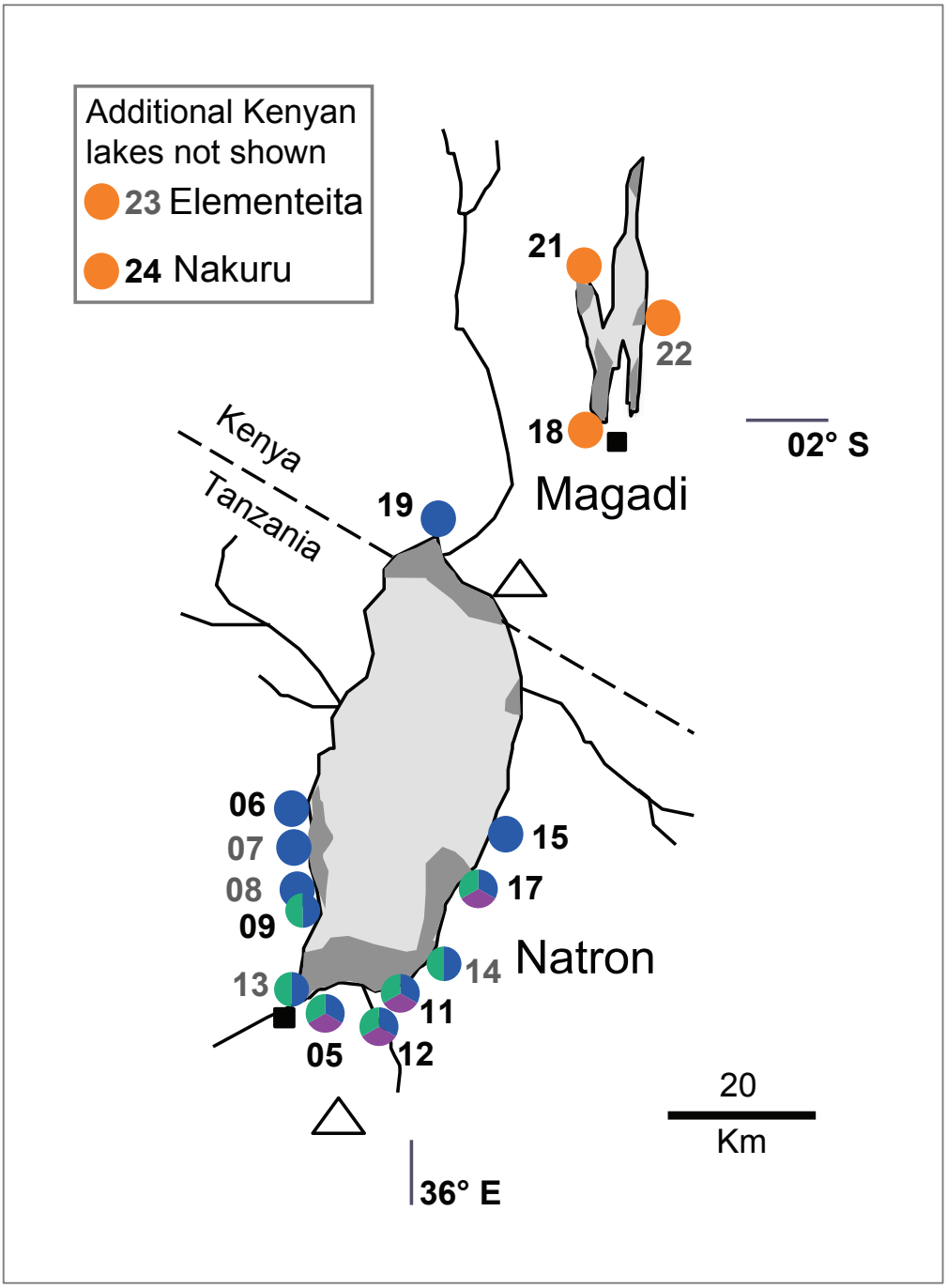


Figure 6.1. Sampling locations for the present study. Morphometric and stable isotope data are available for all sites. Sites labeled in grey text (rather than black) are sampling sites for which only stable isotope and morphometric data area available (no genomic data) and included additionally to the sites in Chapter Five. Colour of the site markers indicates species present at each site. Black squares: type localities for *Alcolapia* species; open triangles: volcanoes. Lake basins are outlined in black, with light grey shading representing trona crust, and dark grey indicating areas of open water (lagoons). Subdivisions of sites 7 (7A and 7B) and 11 (11A, 11B, 11C; not shown individually on map) represent sites of increasing elevation along a single spring.

Stable isotope analysis and stomach contents

Experimental and statistical procedures

All methods, procedures and statistical tests followed those described in chapter five methods.

Inclusion of baseline data

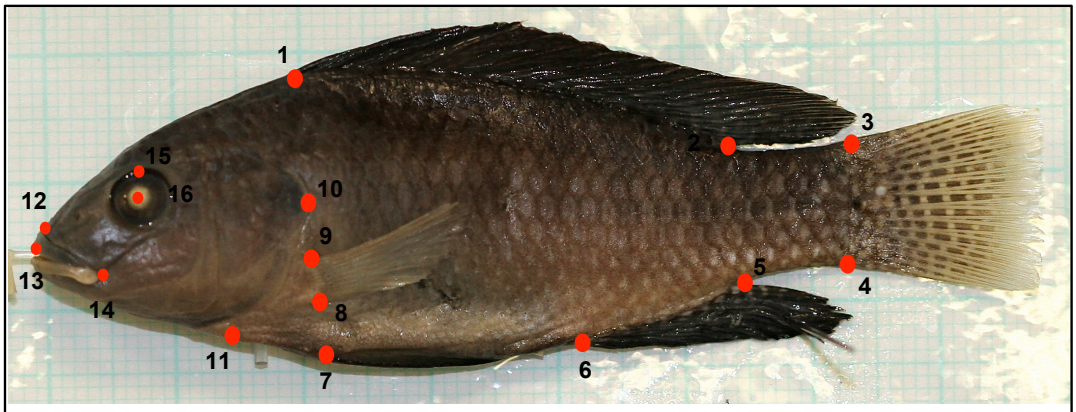
Insufficient baseline (producer and consumer) samples were available from each of the sites to perform baseline-corrections. However, while insufficient to use for correction, algal samples may still provide information about carbon inputs to the food chain, so where available, these samples were included in the analysis (see *Results*).

Sexing by dissection

As external sexing in *Alcolapia* is difficult (except in displaying males and mouthbrooding females), sex was not recorded as standard. However, in order to consider sex ratios, a subset of populations was sexed by dissection for all individuals at three sites in Lake Natron. Sexing was performed by examining gonads via lateral incision or, for those individuals also being dissected for stomach contents, via ventral incision and following removal of the gut. Individuals with undifferentiated gonads, or those where dissection was not possible, were recorded as 'undetermined'.

Geometric morphometrics

Methods and tests for intraspecific variation followed those outlined in chapter five, except for pharyngeal jaw shape, which was only considered for *A. alcalica* intraspecific analyses due to low sample sizes of the other species. An additional GMM shape analysis including only cranial landmarks was undertaken for the *A. alcalica* mouth morphs at site 15, and all three Lake Natron species at sites 5 and 12. The dataset consisted of a subset of the landmarks used for whole body shape variation, including only landmarks 1, 7, 10-16 (Figure 6.2).



1. Anterior insertion of dorsal fin
2. Posterior insertion of dorsal fin
3. Dorsal insertion of caudal fin
4. Ventral base of caudal fin
5. Posterior insertion of anal fin
6. Anterior insertion of anal fin
7. Anterior insertion of pelvic fin
8. Lower insertion of pectoral fin
9. Upper insertion of pectoral fin
10. Posterior extremity of operculum
11. Anteroventral tip of the pectoral girdle
12. Anterior tip of maxilla
13. Lip juncture (anterior snout tip)
14. Posterior point of mouth cleavage
15. Dorsal margin of the eye
16. Centre of the orbit

Figure 6.2. Landmarks included for morphometric analysis of cranial variation. Landmarks described in black text were included for analysis of cranial shape.

Results

Stable isotope analysis

Trophic niche differentiation

A total of 458 individuals were analysed for stable isotope change across all populations (mean: n=16 per population). The values of $\delta^{13}C$ and $\delta^{15}N$ did not co-vary across the whole dataset (Mantel test: 10,000 repetitions, observation=-0.0735, P=0.9996). There was a large variation in both $\delta^{13}C$ (18‰) and $\delta^{15}N$ (15‰) across the dataset, however these differences largely reflected geographic variation, and samples clustered by sampling site rather than species (Figure 6.3), reflecting the impact of baseline levels of isotopes varying between environments. Thus, direct comparisons between sites are not valid based on absolute values, but the total niche area can be compared between sites. Biplots of populations of each species are shown in Appendix Figure 6A.1, again demonstrating the distinct variability in stable isotope ratios between sampling sites. Despite the variation across the dataset, only $\delta^{13}C$ showed variation within sampling sites (Figure 6.4). Bi-plots of SIA variation for additional populations not included in chapter five analysis are shown in Appendix Figure 6A.2.

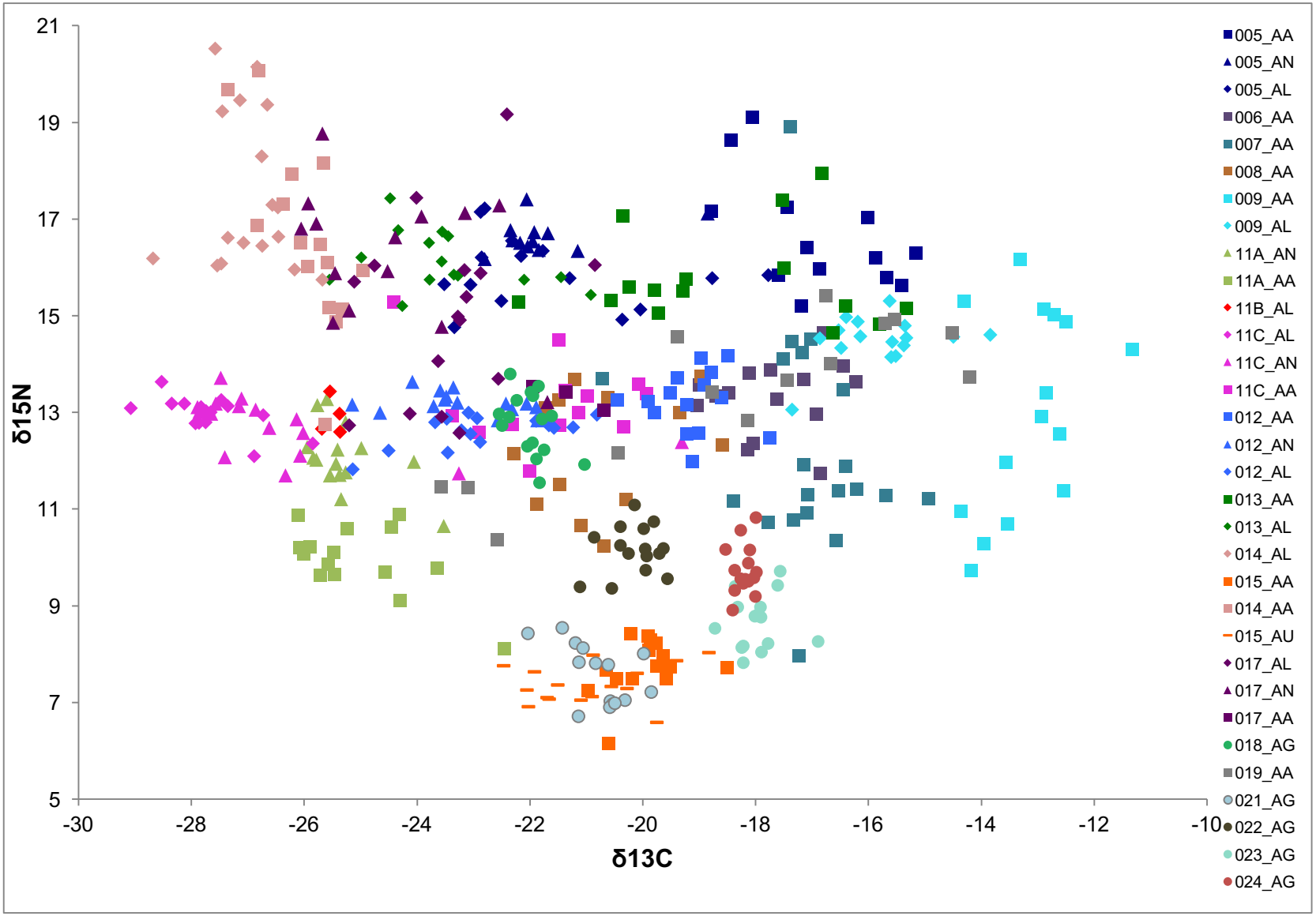


Figure 6.3. Stable isotope ratio results, coloured by population (species/site). Values for $\delta^{13}C$ plotted against $\delta^{15}N$.

AA: *A. alcalica*

AG: *A. grahami*

AL: *A. latilabris*

AN: *A. ndalalani*

AU: *A. alcalica* upturned mouth morph.

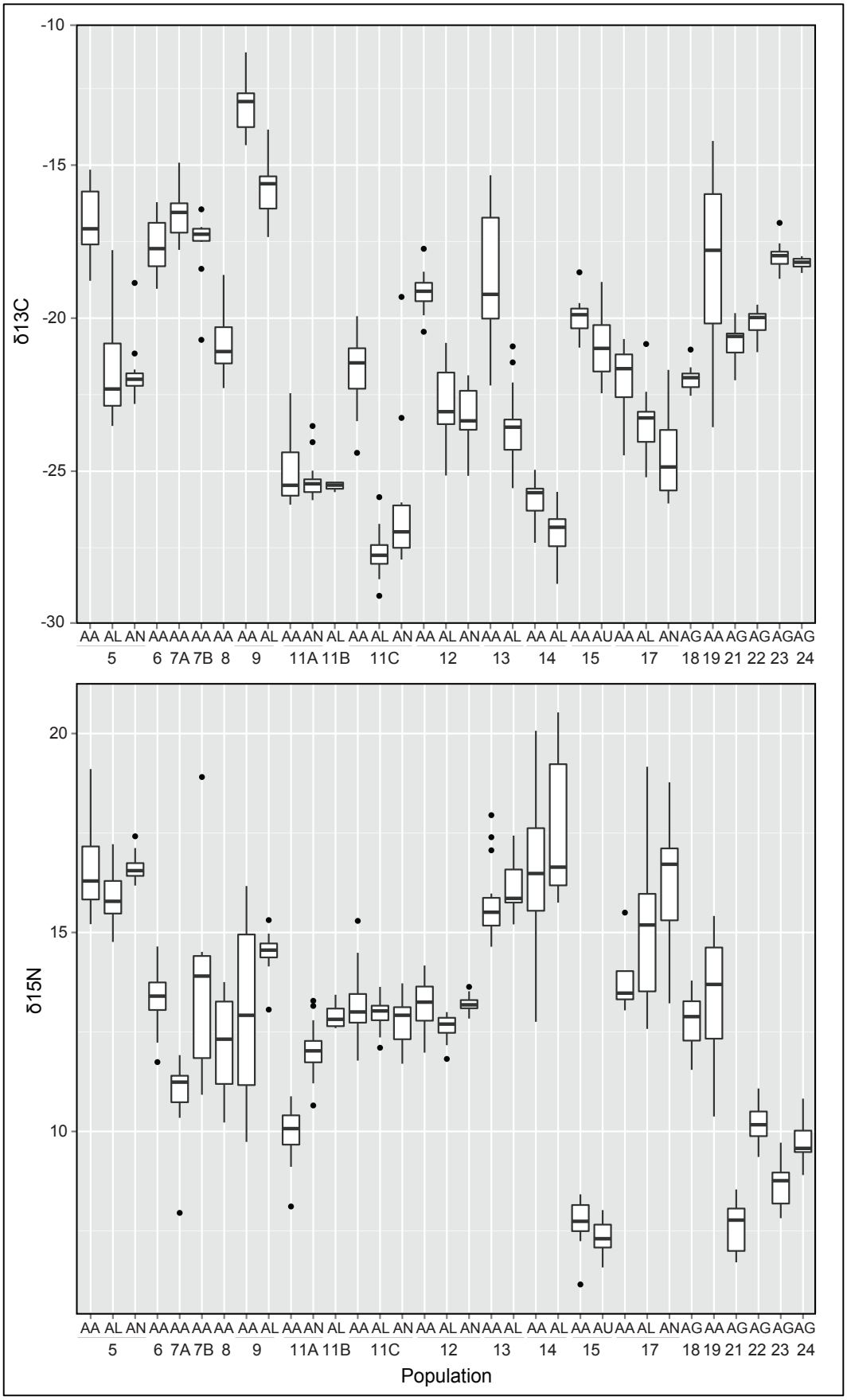


Figure 6.4. Box plots of $\delta^{13}C$ and $\delta^{15}N$ for each population.

A. alcalica is consistently $\delta^{13}C$-enriched relative to other sympatric species.

Trophic niche exploitation in allopatry and geographic variation

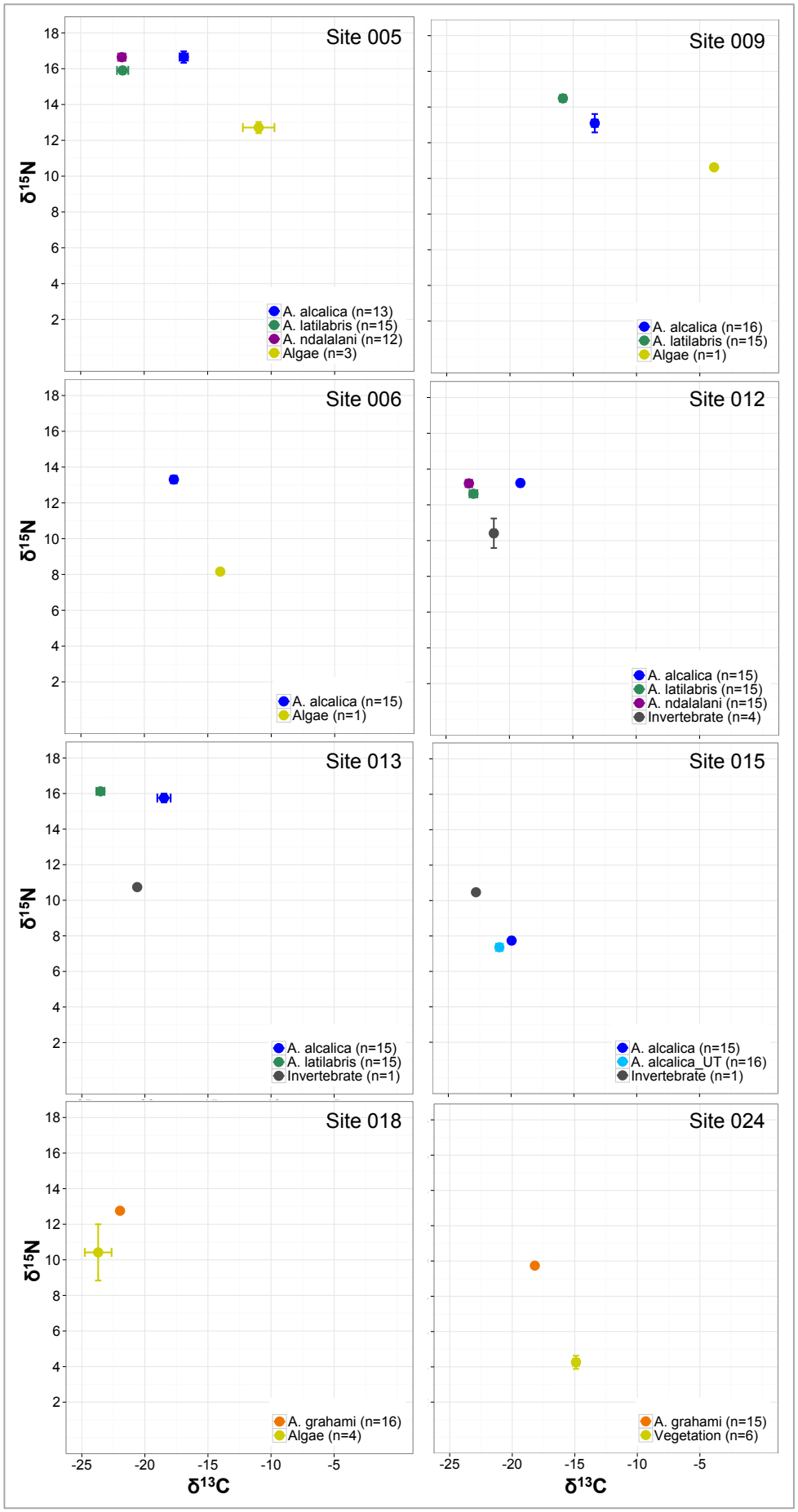
Stable isotope results indicated substantially larger mean niche space occupied by *A. alcalica* and *A. latilabris* compared with *A. grahami* (Table 6.2). While there was a large degree of variability in niche space occupation between populations within Lake Natron species, all *A. grahami* populations exhibited very narrow niche space. *Alcolapia alcalica* occupied significantly broader niche space than *A. grahami* (ANOVA: $F=4.88$; $P<0.01$) and *A. ndalalani* ($F=7.25$; $P<0.01$), but not than *A. latilabris* ($F=2.32$; $P=0.15$). *Alcolapia latilabris* and *A. ndalalani* were not significantly differentiated ($F=0.77$; $P=0.40$). There was no correlation of niche breadth with habitat depth for either *A. alcalica* ($R^2=0.20$, $P=0.27$) or *A. grahami* ($R^2= 0.29$, $P=0.35$). Insufficient data were available for the other species to test a correlation. Although Lake Magadi populations may experience lower allochthonous inputs into the food system due to the absence of inflowing streams present in Lake Natron, which could possibly explain a less variable range of baseline ratios, the populations from larger lakes in which *A. grahami* have been introduced (sites 23 and 24) also display narrow niche space, suggesting this pattern is not population or lake-dependent. *Alcolapia grahami* is the only species recorded in these three lakes (Magadi, Nakuru, Elementeita), so it could also be that due to the lack of competition there has been no selective pressure for niche expansion. Although it is not possible to compare absolute values between sites, plots of all populations per species (given in Appendix Figure 6A.1) show the contracted niche space occupied by *A. grahami* vs. other *Alcolapia* species. Biplots of nitrogen and carbon ratios for the additional populations not presented in chapter five are shown in Appendix Figure 6A.2.

Table 6.2. Niche breadth (ellipse area) for stable isotope values.

Site	*A. alcalica* (n)	*A. latilabris* (n)	*A. ndalalani* (n)	*A. grahami* (n)
5	3.84 (13)	4.16 (15)	1.08 (12)	-
6	2.29 (15)	-	-	-
7	3.23(10)	-	-	-
8	4.63 (13)	-	-	-
9	6.59 (16)	1.43 (15)	-	-
11A	2.06 (15)	-	1.28 (15)	-
11C	4.02 (14)	1.77 (15)	0.74 (15)	-
12	1.38 (15)	1.04 (15)	0.70 (15)	-
13	6.70 (15)	2.73 (15)	-	-
14	2.89 (15)	3.70 (15)	-	-
15	1.06 (AA:15) 1.40 (AU:16)	-	-	-
17	-	6.40 (16)	6.07 (14)	-
19	7.75 (13)			-
18	-	-	-	0.72 (16)
21	-	-	-	1.00 (15)
22	-	-	-	0.74 (15)
23	-	-	-	0.81 (15)
24	-	-	-	0.29 (15)
Mean	**3.68**	**3.03**	**1.97**	**0.71**

Food chain carbon source

For the sites where baseline samples were available, isotopic values are presented in Figure 6.5. The baseline values exhibit considerably different isotopic carbon signatures than the fish samples. However, we would expect fish isotopic ratios to very closely mirror those of the resources on which they are feeding. The discrepancy suggests that the fish are not feeding on algae or invertebrates for which samples were collected, or are feeding on other sources in addition to these. The sample sizes of baseline values are very low, so it is also likely that they do not capture the full range of isotope values for each food source.


Site 005
A. alcalica (n=13)
A. latilabris (n=15)
A. ndalalani (n=12)
Algae (n=3)
Site 009
A. alcalica (n=16)
A. latilabris (n=15)
Algae (n=1)
Site 006
A. alcalica (n=15)
Algae (n=1)
Site 012
A. alcalica (n=15)
A. latilabris (n=15)
A. ndalalani (n=15)
Invertebrate (n=4)
Site 013
A. alcalica (n=15)
A. latilabris (n=15)
Invertebrate (n=1)
Site 015
A. alcalica (n=15)
A. alcalica_UT (n=16)
Invertebrate (n=1)
Site 018
A. grahami (n=16)
Algae (n=4)
Site 024
A. grahami (n=15)
Vegetation (n=6)
δ15N
δ13C

Stomach contents and gut length

A total of 178 individuals were measured for gut length and 185 analysed for stomach contents (92% individuals were included for both analyses, with 8% included in only one analysis). Gut length varied between populations (Figure 6.6A). For all Lake Natron species, individuals from site 11 had lower gut/body length ratio than for intraspecific comparisons at sites 5 and 12. Stomach contents of *A. alcalica* varied between the southern and northern sites, with the southern populations (sites 5, 11, 12) having a large proportion of contents accounted for by cellulose (32-49%), while northern populations (sites 15 and 19) consumed a much lower proportion of cellulose (0-6%) (Figure 6.6B). *Alcolapia alcalica* individuals from site 19 had a much higher proportion of grit/sand than conspecifics from other populations. For *A. ndalalani*, *A. latilabris* and *A. grahami*, diet consisted of a high proportion of algae across all populations (38-81%), and *A. latilabris* had a high proportion of grit/sand at all sites (37-62%).

Figure 6.5 (opposite page). Stable isotope ratios (Mean +/- SEM) per site, including baseline (vegetation and invertebrate) values.

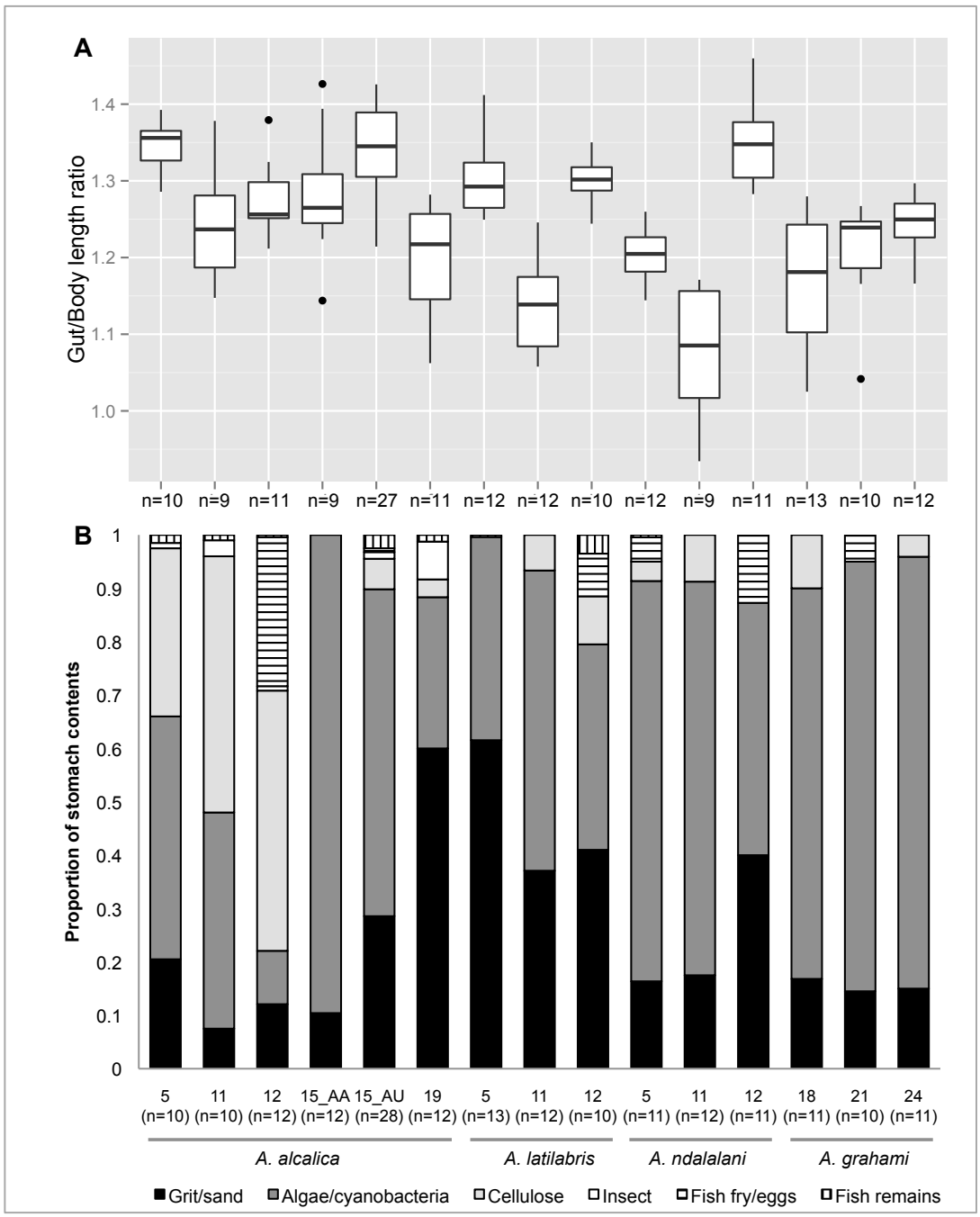


Figure 6.6. Gut length and stomach contents by population.

Species and population (site) designation are given on the x-axis of the lower graph. Sample numbers are given separately for each graph as these differed between analyses. A) Relative gut length (log_{10} gut/body length); B) Stomach contents analysis by population.

The differentiation between northern and southern *A. alcalica* populations was reflected in Schoener's index of overlap, where few populations exhibited overlap, while there was substantial (>0.6) overlap between all intraspecific comparisons for *A. latilabris*, *A. ndalalani* and *A. grahami* (Figure 6.7). In the *A. alcalica* comparison,

the northern sites (15 and 19) appeared most ecologically disparate from southern sites based on Schoener's index (Figure 6.7). Furthermore, the morphs (upturned-mouth and terminal mouth) within site 15 showed considerable overlap. While this is in line with the stable isotope results showing overlap in niche space (chapter five, Figure 5.5), it had been expected that the morphs would be ecologically differentiated in diet based on differences in trophic morphology.

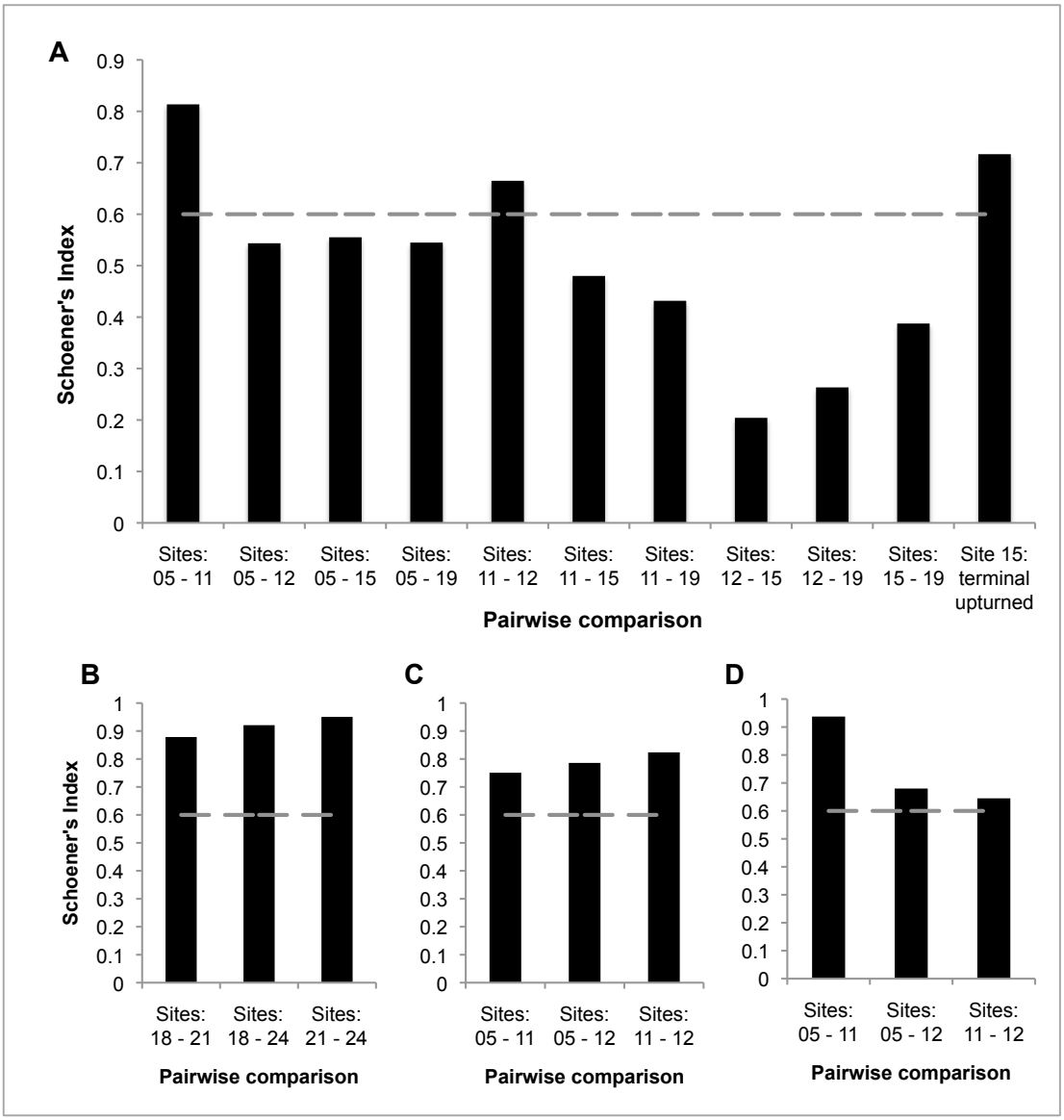


Figure 6.7. Schoener's Index of dietary overlap for intraspecific comparisons.
A) *A. alcalica*; B) *A. grahami;* C*) A. latilabris*; D) *A. ndalalani*. Grey dashed line indicates proportion at which overlap is considered significant.

Sex ratios

During dissection for stomach contents analysis, it was apparent that the sex ratios in one of the populations (site 15) appeared to be skewed. To investigate this in more detail, individuals from a subset of populations were dissected to determine sex. All individuals from sites 5, 12, and 15 were dissected. The results confirmed the skewed sex ratio observation, with all terminal mouth morphs at site 15 being male, and all upturned-mouth morphs being female (excluding individuals of undetermined sex) (Figure 6.8).

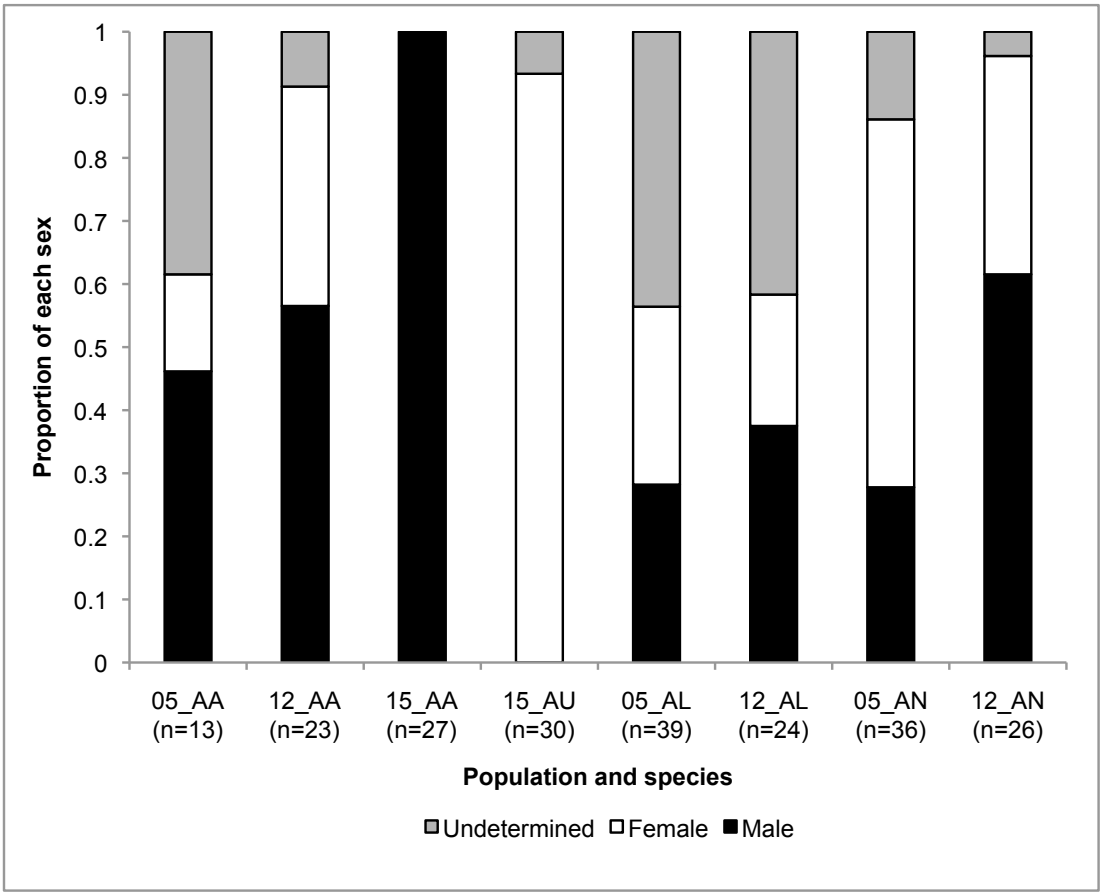


Figure 6.8. Proportion of specimens by sex for three sampling sites (5, 12, 15).
AA: *A. alcalica*, AU: *A. alcalica* upturned-mouth morph; AL: *A. latilabris*; AN: *A. ndalalani*.

It was also notable that nearly all samples from site 15 exhibited differentiated gonads (including males as small as 30mm standard length). Although the moderately small sample sizes preclude definitive conclusions being drawn from these data, it would suggest that either the terminal and upturned-mouth morphotypes at site 15 represent intraspecific sexual dimorphism rather than

ecotypes, or that for some reason sampling at site 15 was biased by sex for each morphotype.

Geometric Morphometrics

Variation of body shape – all populations

A total of 734 individuals were included for the geometric morphometric analysis of body shape in the full analysis including all populations. As species-level divergence was investigated in the previous analysis (in chapter five), this full analysis only considered population-level divergence within each of the species. Regression of Procrustes coordinates against centroid size for all 734 individuals was significant (10,000 permutations; $P<0.0001$) and size accounted for 3.26% of the variation within the dataset, thereafter size-corrected results (regression residuals) were used for all downstream analyses. In a PCA on the Procrustes-fitted and sized-corrected residuals across the entire dataset, the first three PCs accounted for 65% of the variation within the dataset. Eigenvalues and PCA plot (PC1 and PC2) for the analysis including all specimens are given in Appendix Figure 6A.3 and Appendix Table 6A.2. Principal components analysis across the entire dataset gave similar results to that for the reduced RAD-dataset, but contained too many populations to analyse concurrently, and further analyses were not conducted on the entire dataset. Intraspecific variation in body shape for populations in each species analysed individually is given in Figures 6.9-6.12. Where the dataset was divided by species, a new Procrustes alignment and regression for size-correction was performed for each dataset.

For intraspecific analysis, site 17 was excluded from the *A. alcalica* comparison due to small sample size and inconclusive species assignation. Principal component analysis of the remaining *A. alcalica* populations revealed variation in mouth orientation (terminal or upturned) and body depth, although most populations overlapped in morphospace to some extent (Figure 6.9). Canonical variate analysis showed a clearer pattern of differentiation; with the majority of pairwise comparisons using Procrustes distances from CVA significant (Table 6.3) after Bonferroni correction for multiple tests. Populations 11C, 14 and 15 (AU) appeared as particularly differentiated from other populations in the CVA plot (Figure 6.9).

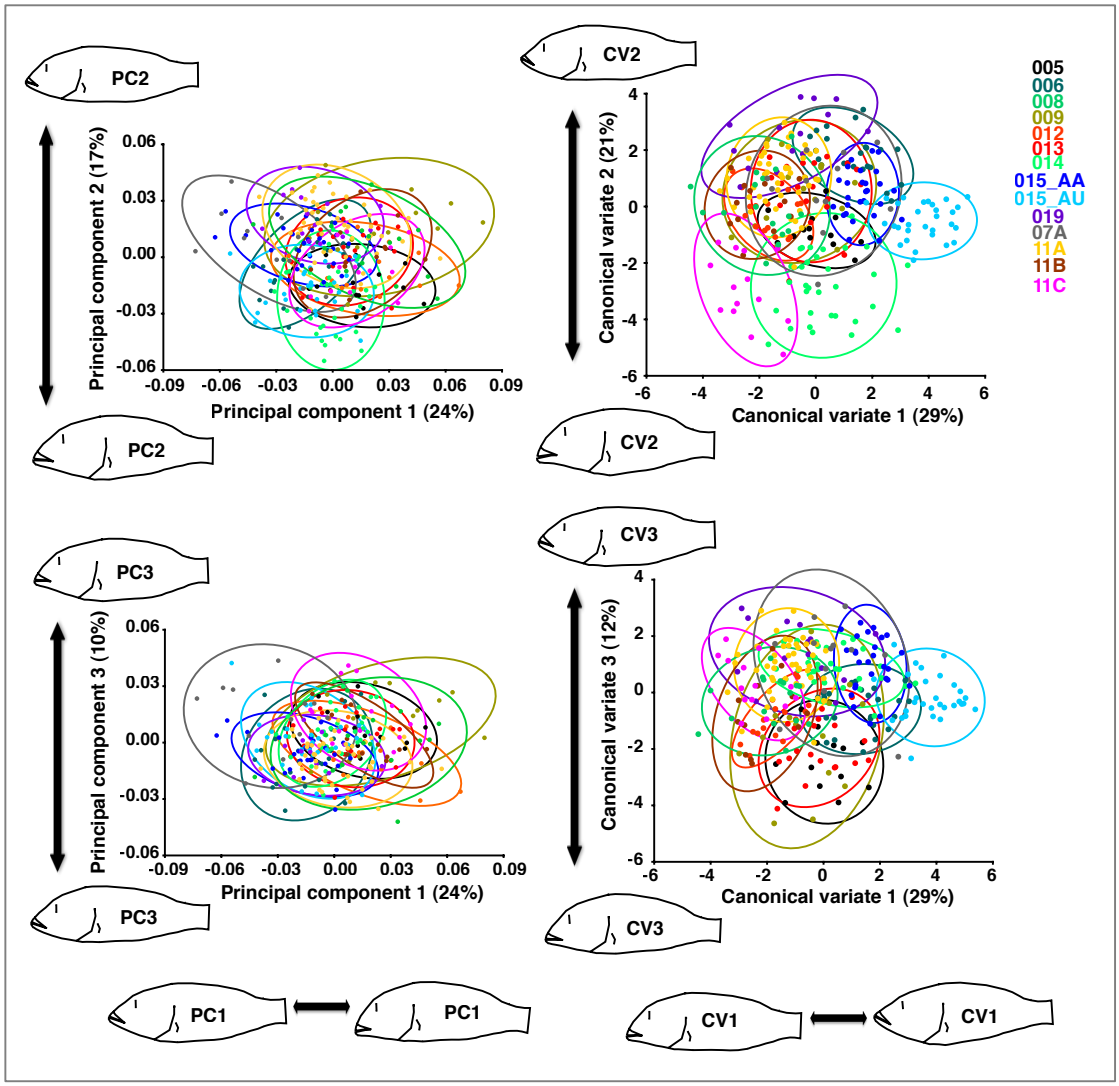


Figure 6.9. Body shape analysis (PCA and CVA) for *A. alcalica* populations.

Table 6.3. Pairwise Procrustes distances between *A. alcalica* populations.

Site	5	6	8	9	12	13	14	15AA	015A	19	07	11A	11B
5 (n=13)	-												
6 (n=24)	0.04	-											
8 (n=16)	0.03	0.04	-										
9 (n=14)	0.04	0.05	0.03	-									
12 (n=16)	0.03	0.04	0.03	0.04	-								
13 (n=19)	0.03	0.03	0.03	0.03	0.03	-							
14 (n=33)	0.03	0.04	0.04	0.05	0.04	0.04	-						
15AA (n=27)	0.05	0.03	0.04	0.05	0.05	0.04	0.04	-					
15AU (n=30)	0.04	0.02	0.05	0.05	0.05	0.04	0.03	0.03	-				
19 (n=14)	0.05	0.03	0.04	0.04	0.05	0.04	0.05	0.03	0.05	-			
7A (n=12)	0.06	0.04	0.06	0.06	0.06	0.05	0.05	0.04	0.04	0.04	-		
11A (n=47)	0.04	0.03	0.02	0.03	0.03	0.03	0.04	0.03	0.04	0.02	0.05	-	
11B (n=12)	0.03	0.04	0.02	0.03	0.03	0.02	0.04	0.04	0.05	0.04	0.06	0.03	-
11C (n=14)	0.03	0.05	0.03	0.04	0.04	0.03	0.03	0.05	0.05	0.05	0.06	0.04	0.04

Unshaded cells represent comparisons significant following Bonferroni correction (Grey shading indicates non-significance).

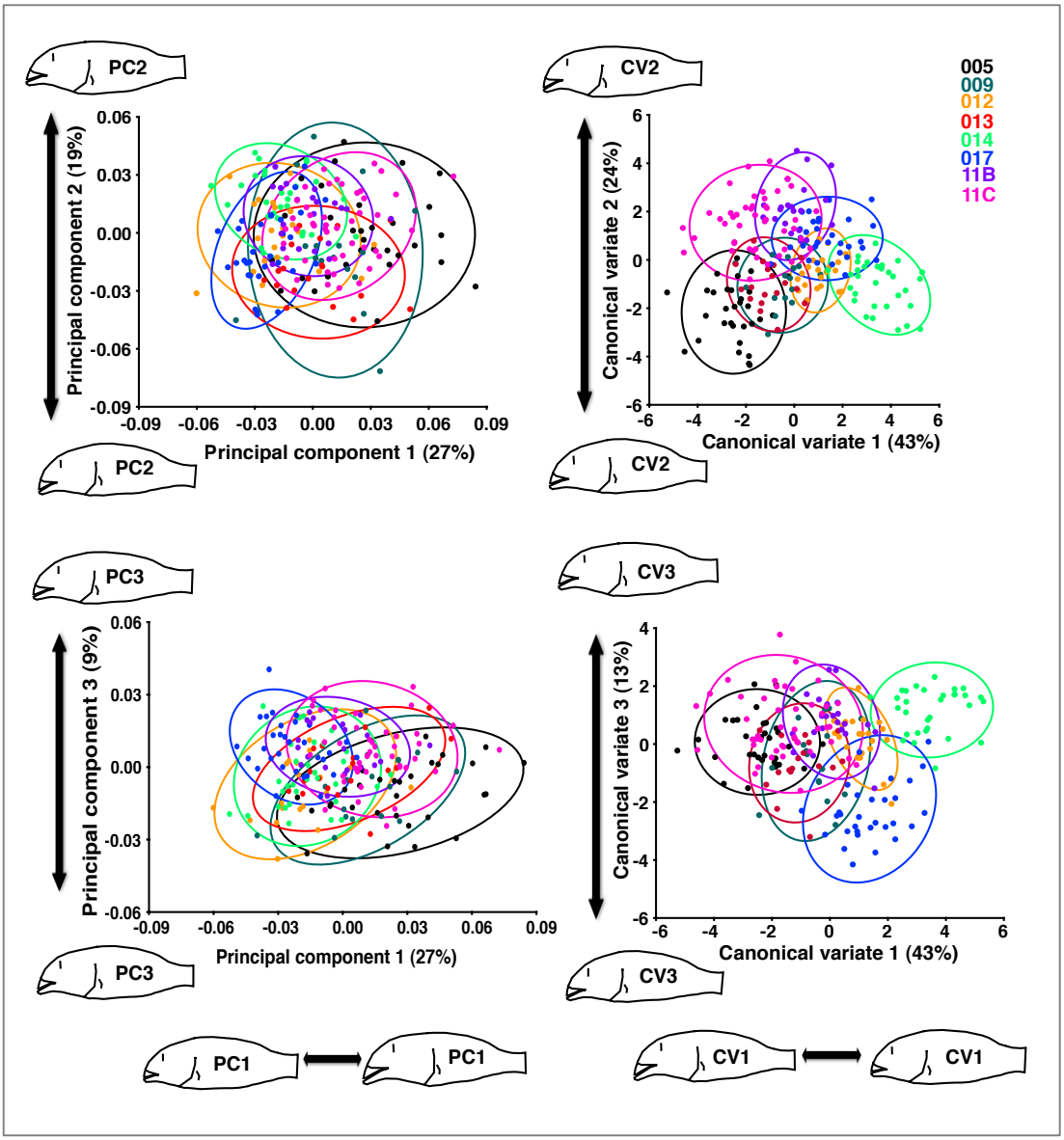


Figure 6.10. Body shape analysis (PCA and CVA) for *A. latilabris* populations.

Table 6.4. Pairwise Procrustes distances between *A. latilabris* populations.

	5	9	12	13	14	17	11B
5 (n=31)							
9 (n=16)	0.03						
12 (n=22)	0.05	0.03					
13 (n=18)	0.04	0.03	0.03				
14 (n=31)	0.05	0.04	0.02	0.05			
17 (n=31)	0.06	0.04	0.02	0.04	0.04		
11B (n=21)	0.04	0.03	0.03	0.04	0.03	0.03	
11C (n=45)	0.03	0.03	0.04	0.03	0.04	0.04	0.02

Unshaded cells represent comparisons significant following Bonferroni correction.

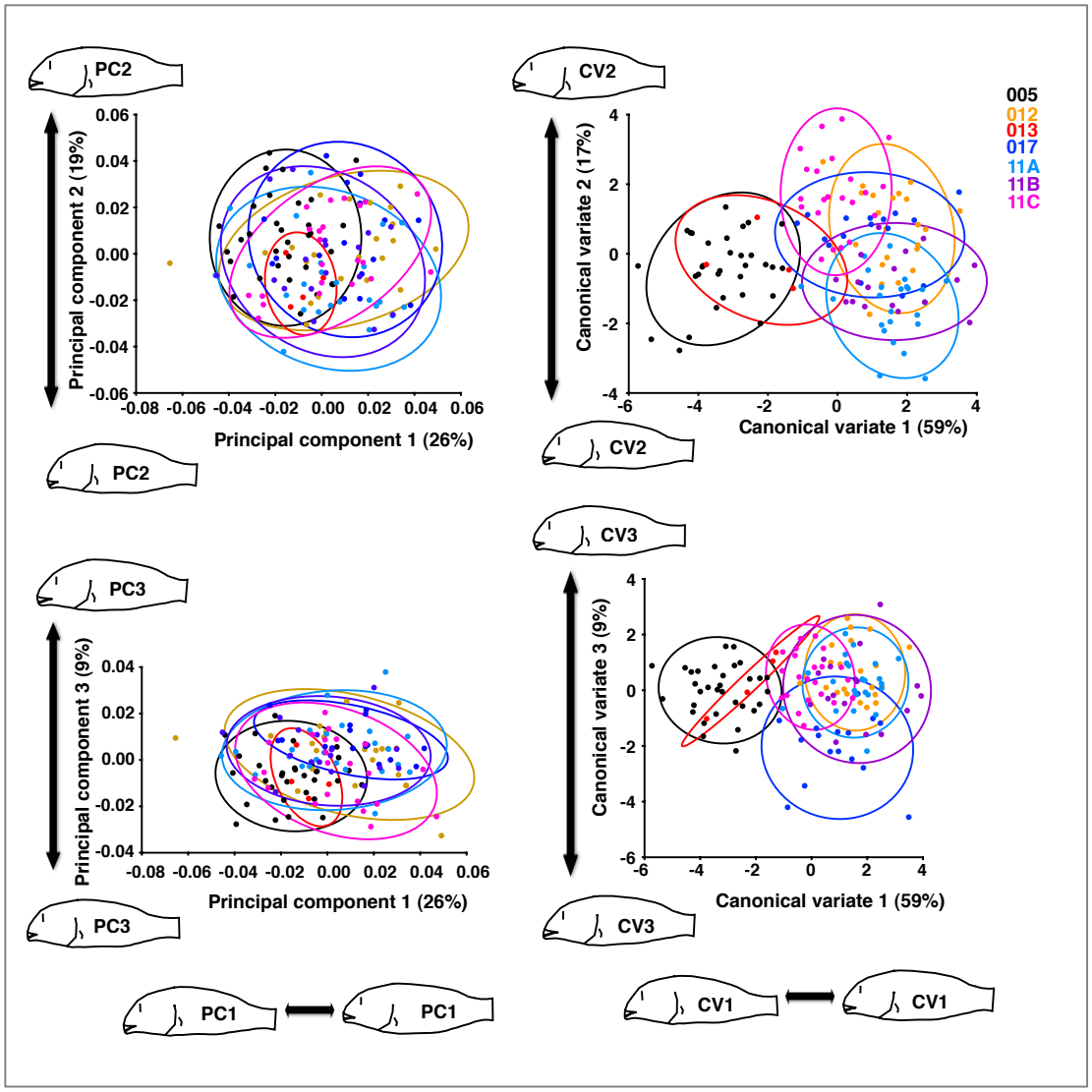


Figure 6.11. Body shape analysis (PCA and CVA) for *A. ndalalani* populations.

Table 6.5. Pairwise Procrustes distances between *A. ndalalani* populations.

	5	12	13	17	11A	11B
5 (n=32)						
12 (n=22)	0.03					
13 (n=4)	0.02	0.03				
17 (n=16)	0.03	0.02	0.04			
11A (n=24)	0.03	0.02	0.03	0.02		
11B (n=17)	0.03	0.02	0.03	0.02	0.01	
11C (n=22)	0.03	0.02	0.03	0.02	0.02	0.02

Unshaded cells represent comparisons significant following Bonferroni correction.

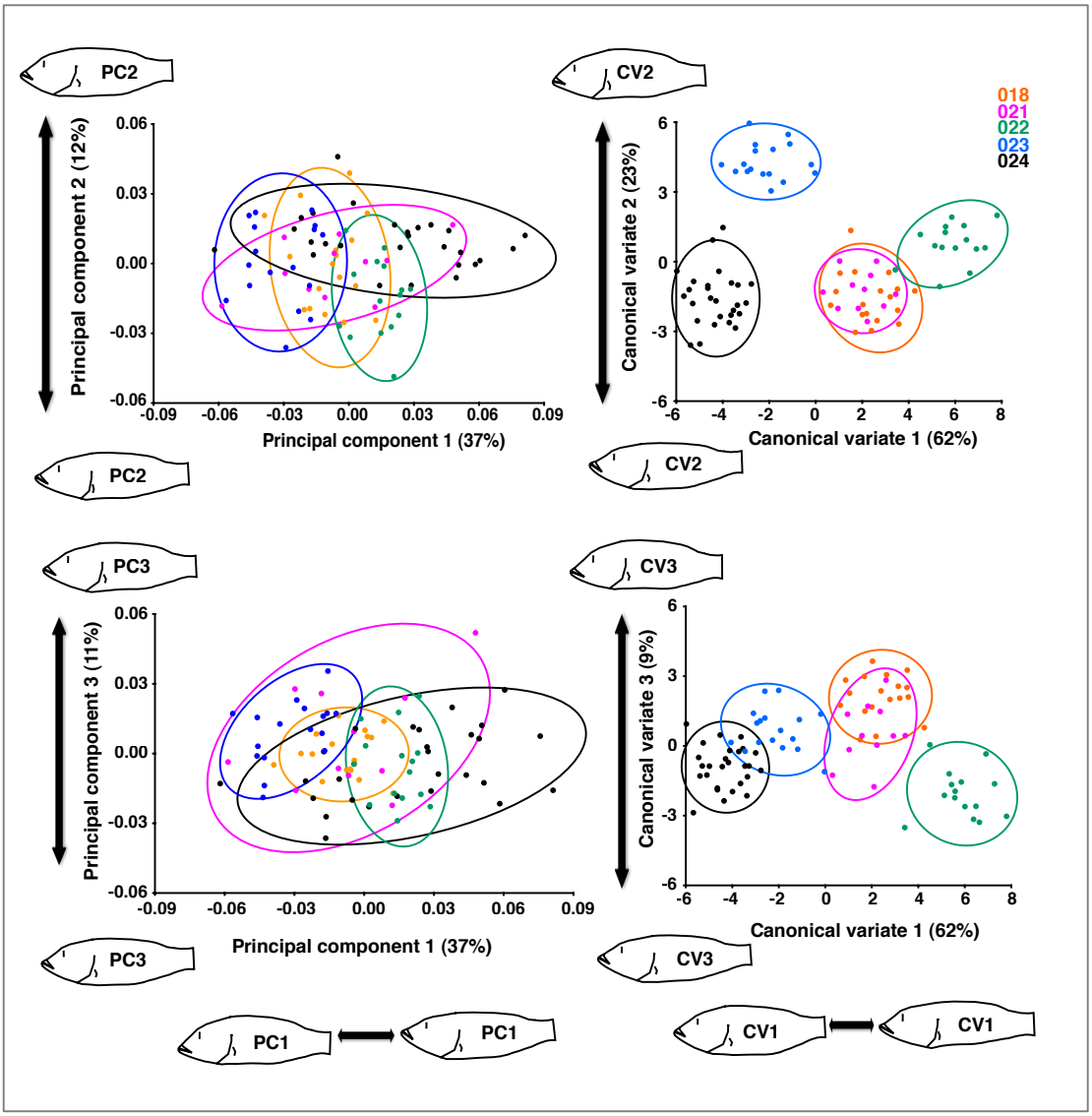


Figure 6.12. Body shape analysis (PCA and CVA) for *A. grahami* populations.

Table 6.6. Pairwise Procrustes distances between *A. grahami* populations.

Site	18	21	22	23
18 (n=18)				
21 (n=11)	0.02			
22 (n=15)	0.04	0.04		
23 (n=17)	0.04	0.04	0.06	
24 (n=27)	0.04	0.04	0.05	0.06

Unshaded cells represent comparisons significant following Bonferroni correction.

Significant differences were also seen at the population level within the other three *Alcolapia* species as follows (p-values from 10,000 rounds permutation tests, Procrustes distances, P<0.05 following Bonferroni correction): *A. latilabris*: all populations significantly differentiated except site 5 vs. site 9 and site 9 vs. 13 (Table 6.4); *A. ndalalani*: significant differentiation only for comparisons of site 5 to other sites (Table 6.5); *A. grahami*: all populations significantly differentiated except site 18 vs. site 21 (Table 6.6). Removing the introduced populations of *A. grahami* (sites 23 and 24) and running CVA on only the Lake Magadi populations produced similar results, with site 22 differentiated from both site 18 and 21 (P<0.001), but 18 and 21 not differentiated from each other (P=0.18). Plots of PCA and CVA for Magadi *A. grahami* populations are given in Appendix Figure 6A.4.

Despite the significant differences between populations, most Lake Natron species populations overlapped in PCA. Canonical variate analysis separated populations to some extent, with several populations appearing differentiated: site 11C (all species), site 014 (*A. alcalica* and *A. latilabris*), and site 005 (only within *A. ndalalani* comparisons). Populations of *A. grahami* were significantly more differentiated than the Lake Natron species (Figure 6.12), and although all Lake Magadi populations overlapped in morphospace both of the other (introduced) populations at Lake Elementeita (23) and Lake Nakuru (24) were substantially differentiated from other populations. Populations that were differentiated across multiple species in Lake Natron (11C and 014) were reanalysed in a separate sub-analysis for each site to investigate these patterns further.

Principal components analysis of all species at site 11C revealed some overlap between *A. alcalica* and *A. ndalalani* at this site, but all three species were substantially differentiated (Figure 6.13) and significantly differentiated in CVA analysis (all pairwise comparisons, P<0.0001). However, separate analysis of site 014 revealed substantial overlap of *A. alcalica* and *A. latilabris*, such that there appeared to be no differentiation between the two species at this site (Figure 6.14). For comparison, this sub-analysis was combined with another site at which *A. alcalica* and *A. latilabris* occurred alone together (site 009), and PCA on this combined dataset revealed that site 014 individuals overlapped together in morphospace between *A. alcalica* and *A. latilabris* from site 009 (Figure 6.14). When analysed by CVA, 014 *A. alcalica* and *A. latilabris* overlapped entirely while *A. alcalica* and *A. latilabris* from site 009 were completely differentiated, and all pairwise comparisons were significant in permutation tests (P<0.0001) except the *A. alcalica/A. latilabris* site 014 comparisons which was non-significant (P=0.82).

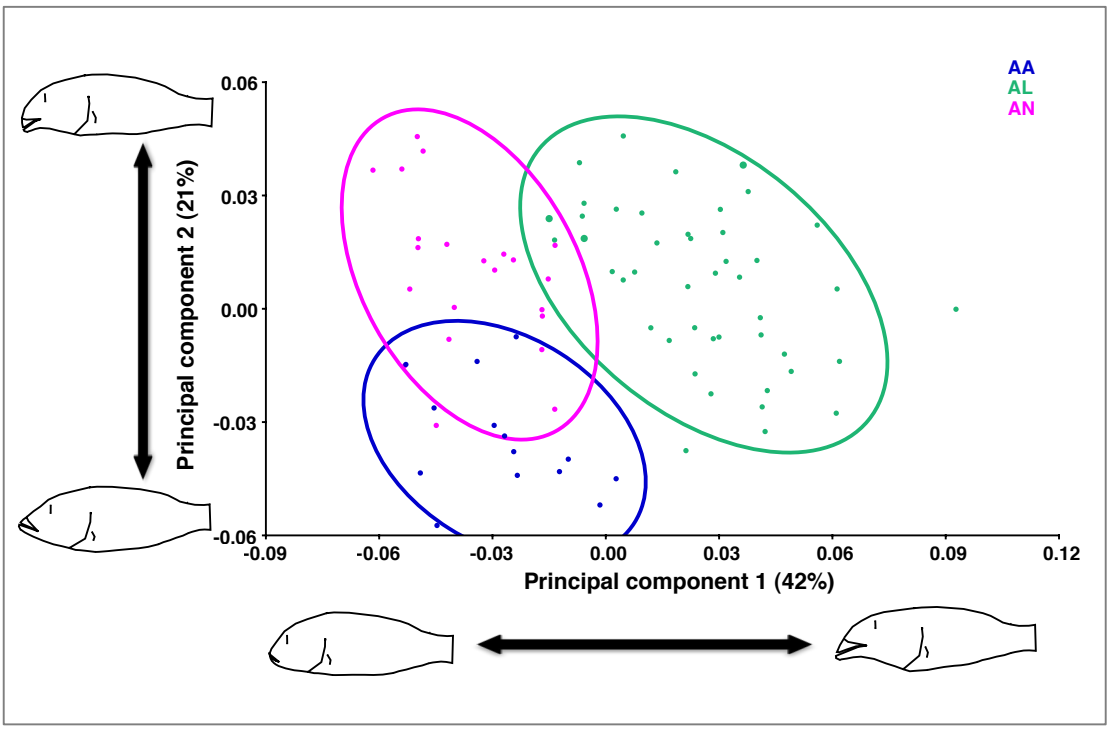


Figure 6.13. Principal components analysis for sympatric species at site 11C.

All comparisons were significantly differentiated in CVA (not shown).

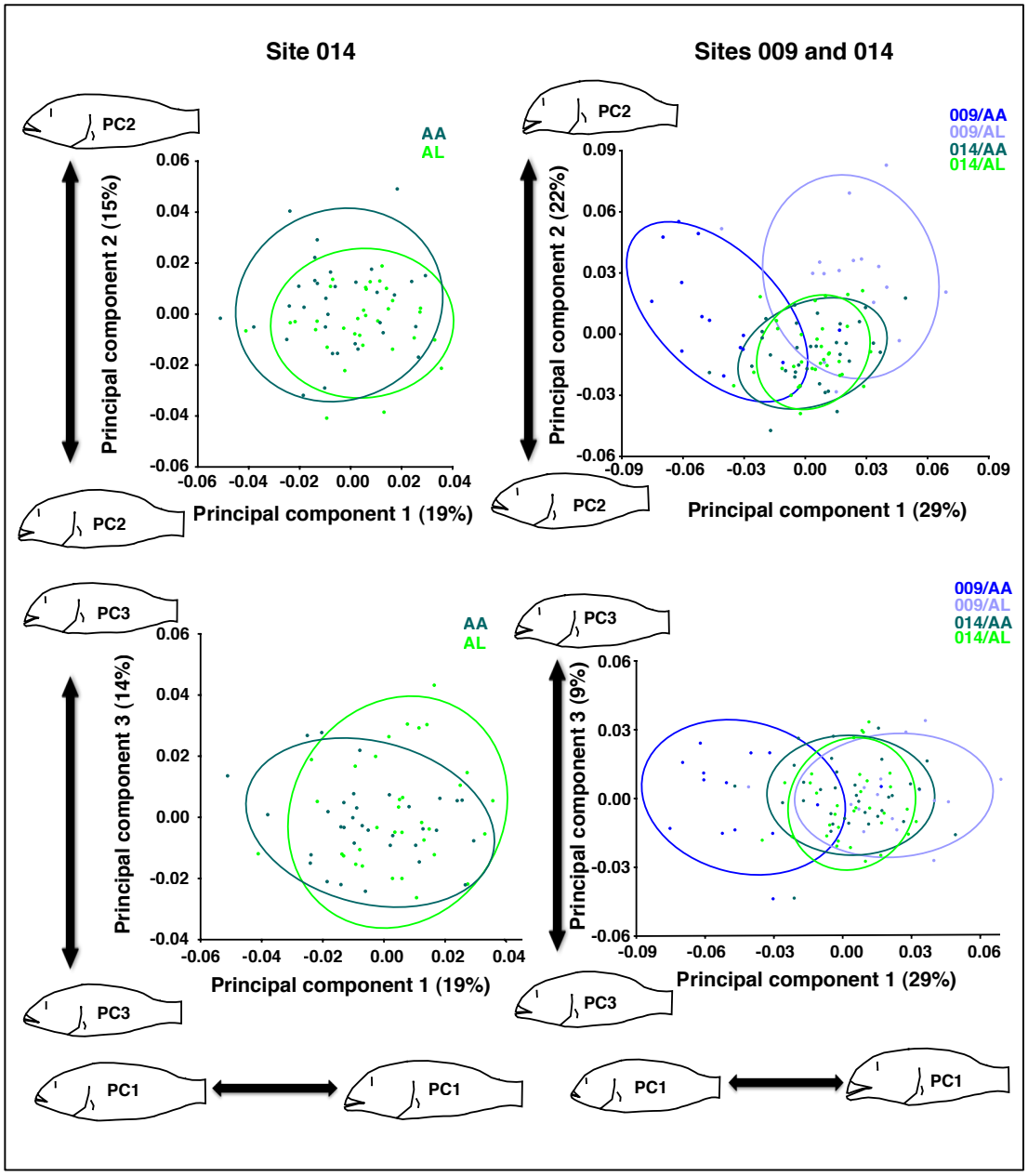


Figure 6.14. Principal components analysis of site 14 individuals.

Alcolapia alcalica, AA, and *A. latilabris*, AL (left-hand column), and a combined analysis of PCA on individuals from sites 009 and 014 (right-hand column) demonstrating substantial overlap between *A. alcalica* and *A. latilabris* at site 014 in comparison to site 009.

That species at site 14 were not clearly differentiated was observed in initial species ID following keys, as the *A. latilabris* individuals did not exhibit the characteristic inferior mouth position typical of the species, and *A. alcalica* individuals appeared to have slightly thickened lips. As these traits complicated species ID, initial species labelling relied on width of mouth rather than typical species traits. However, based on the GMM results (Figure 6.14), it would appear that this feature is not sufficient to

differentiate species (although lateral measures such as width were not accounted for in the landmark distribution) indicating that the individuals exist along a continuum at this site. Photographs of the two morphotypes are provided in Appendix Figure 6A.5. The original species designation did not consider size differences, but mean SL was significantly greater (ANOVA: $F=53.77$, $P<0.0001$) in individuals designated *A. alcalica* (mean: 50.7 mm; range: 29-76 mm) than those designated *A. latilabris* (mean: 37.1 mm; range: 26-49 mm). The overlap in range of these sizes suggests that the designations do not simply describe allometric differentiation of a single morphotype.

Differentiation by sex in GMM

Given the finding that morphotypes from site 15 (terminal and upturned-mouth morphs) differentiated by sex (Figure 6.8), sex and morphology was re-examined for the three Lake Natron populations where all individuals were sexed (5, 12, 15) to investigate if *Alcolapia* display sexual dimorphism. Principal component analysis of all populations combined, and coloured by species (Figure 6.15A) or population (Figure 6.15B) revealed that sexes cluster by species and population. Separate PCA by species revealed that sexes were not differentiated in morphology for *A. alcalica*, *A. latilabris* and *A. ndalalani* at sites 5 and 12 (sites combined; Figure 6.15C-E). Discriminant function analysis also did not find significant differences between sexes at these sites. However, PCA on *A. alcalica* individuals from site 15 revealed that the terminal and upturned morphs (sexes) were largely differentiated, with minimal overlap in morphospace (Figure 6.14F), and DFA between the sexes found a significant shape difference ($P=0.0001$).

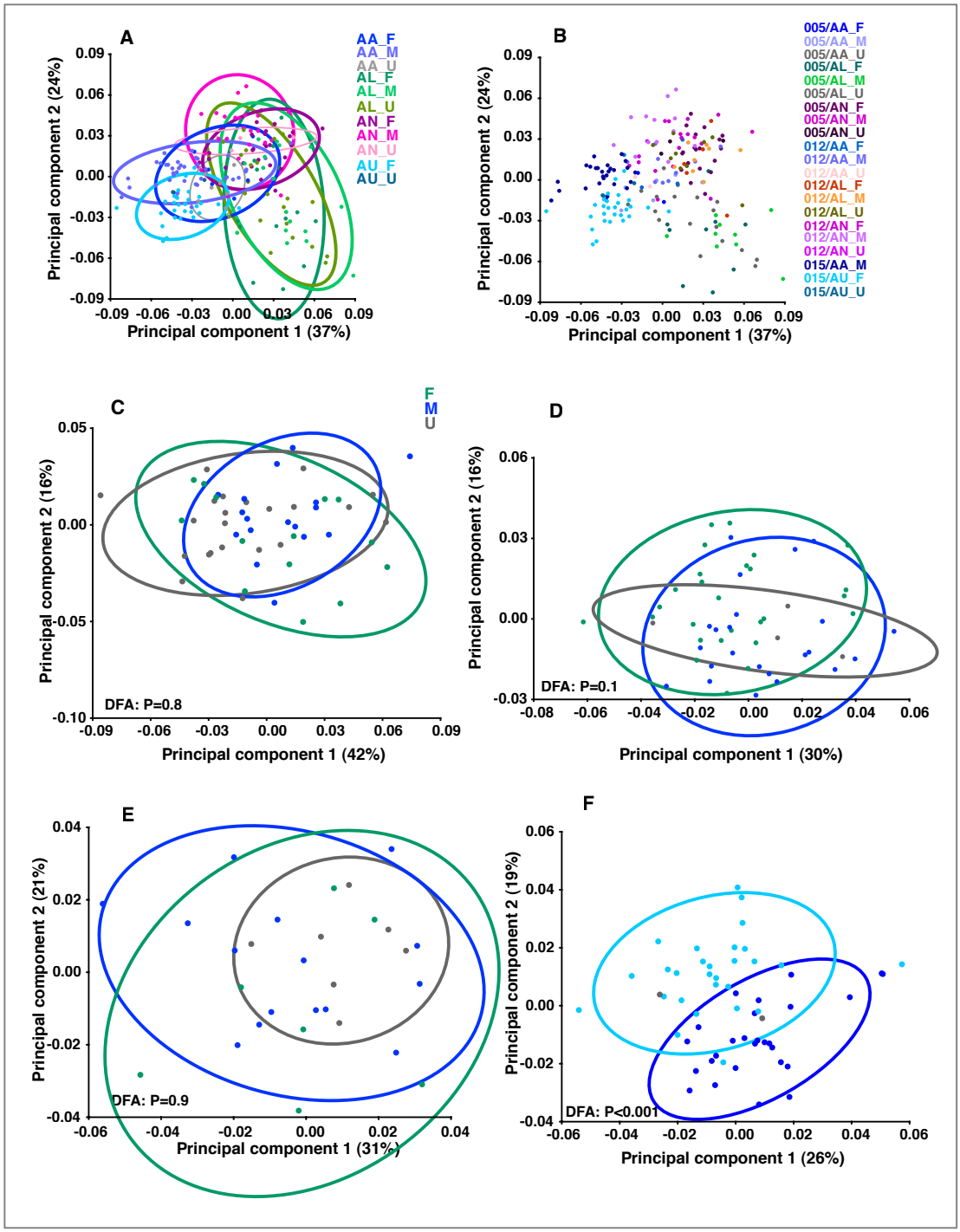


Figure 6.15. Body shape PCA analysis for populations 5, 12, 15, separated by sex.

A) PCA coloured by species/sex. B) Coloured by population/sex. C-E) Coloured by sex: female, green; male, blue; undetermined, grey. C) *A. latilabris*; D) *A. ndalalani*; E) *A. alcalica*; F) *A. alcalica* site 15, terminal morph (all male), blue; upturned mouth morph (female), light blue; undetermined, grey (2 samples, no ellipse). P-values are included in plots C-F from discriminant function analysis for male-female comparisons.

As the morphs from site 15 had originally been identified based on mouth morphology, and mouth angle has previously been shown to exhibit sexual

dimorphism in another cichlid species (Lake Tanganyika *A. burtoni*; Theis *et al.* 2014), a separate GMM analysis was performed on cranial landmarks alone to ensure that homogeneity in post-cranial body shape did not obscure a signal in mouth shape or orientation. Landmarks were retained that described the mouth, orbit, operculum, and insertion points of dorsal and pelvic fins (Figure 6.2). Each of the four datasets described above (three species at sites 5 and 12; morphs at site 15) was analysed separately. Discriminant function analysis revealed minimal differentiation between sexes for *A. latilabris*, with a slightly narrower head in females, although shape differences were not significant (Figure 6.16A). Shape differences between sexes for *A. ndalalani* were marginally insignificant (P=0.054), and which described a narrower, longer head with more rounded dorsal profile and more downward-turned mouth in females (Figure 6.16B). Shape changes for southern *A. alcalica* followed a similar pattern, with females having a more slender and longer head but with no difference in mouth angle (although these differences were not significant in DFA; Figure 6.16C). Finally, DFA between the *A. alcalica* terminal-mouth (male) and upturned-mouth (female) morphs at site 15 revealed significant shape changes in mouth angle and head length (Figure 6.16D).

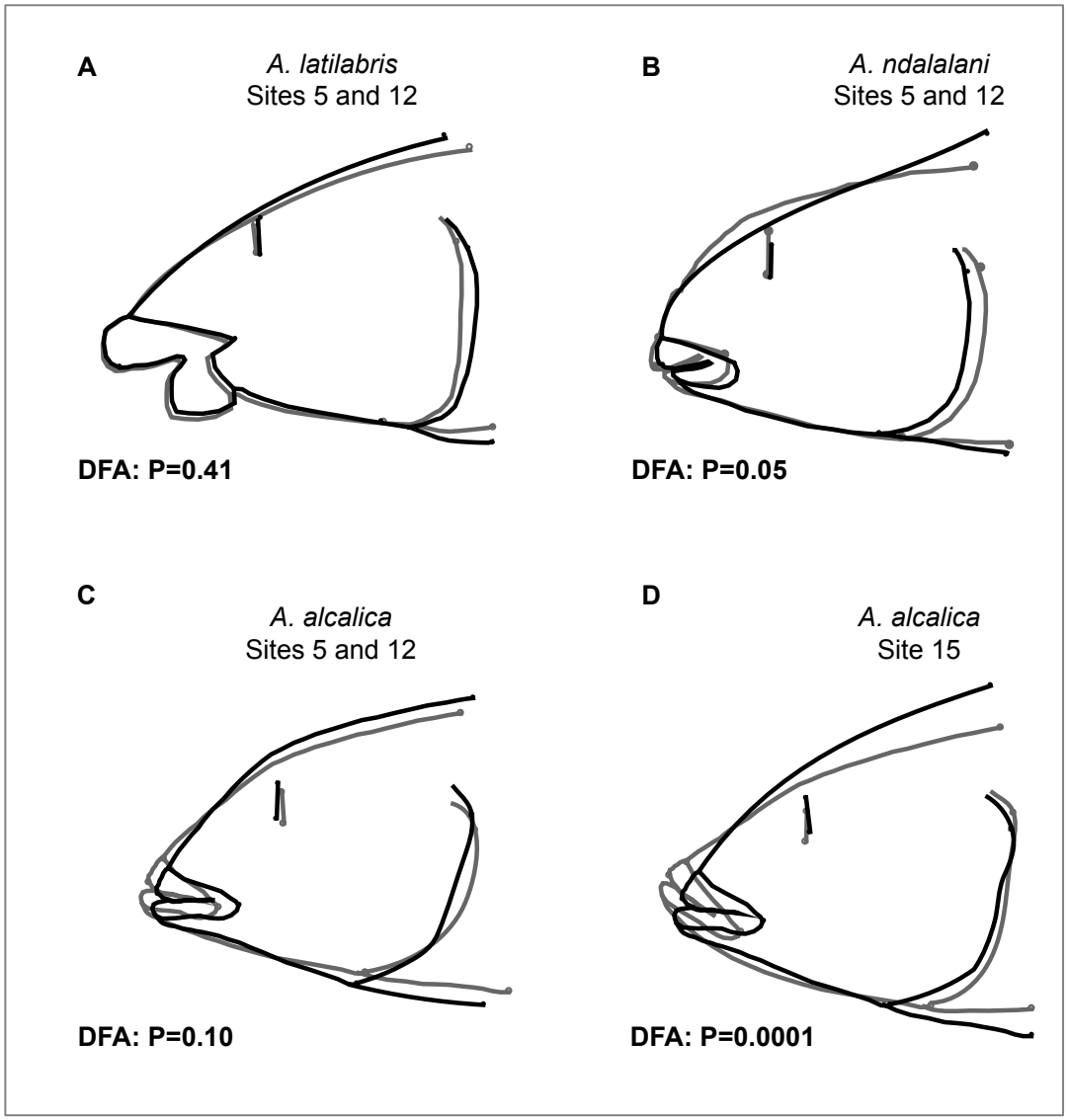


Figure 6.16. Head shape changes for populations 5, 12, 15, separated by sex.
Shape changes from DFA of male-female comparisons for each species at sites 5 and 12, and *A. alcalica* morph comparison at site 15; black, male; grey, female. P-value for each comparison is Mahalanobis T-square value from DFA. Shape changes are scaled by a factor of 3. Plots of PCA for the cranial landmark datasets are given in Appendix Figure 6A.6.

Variation of lower pharyngeal jaw shape – *A. alcalica* populations

Geometric morphometric analysis of four *A. alcalica* populations for LPJ shape (mean n=10 specimens per population) revealed only very subtle differences between populations (Figure 6.17). While principal components analysis did not separate populations, CVA exhibited significant differentiation between southern (5, 12) and northern sites (15, 19) (Table 6.7). The shape differences were mainly described by the height/width of the jaw (PC1 and CV1) and the curvature or the

horns and tooth plate (PC2 and CV2). This may reflect differences in the diets between populations with individuals from southern sites consuming a considerably higher proportion of cellulose (Figure 6.6), although increased sample numbers would be needed to examine these fine-scale population differences between sites.

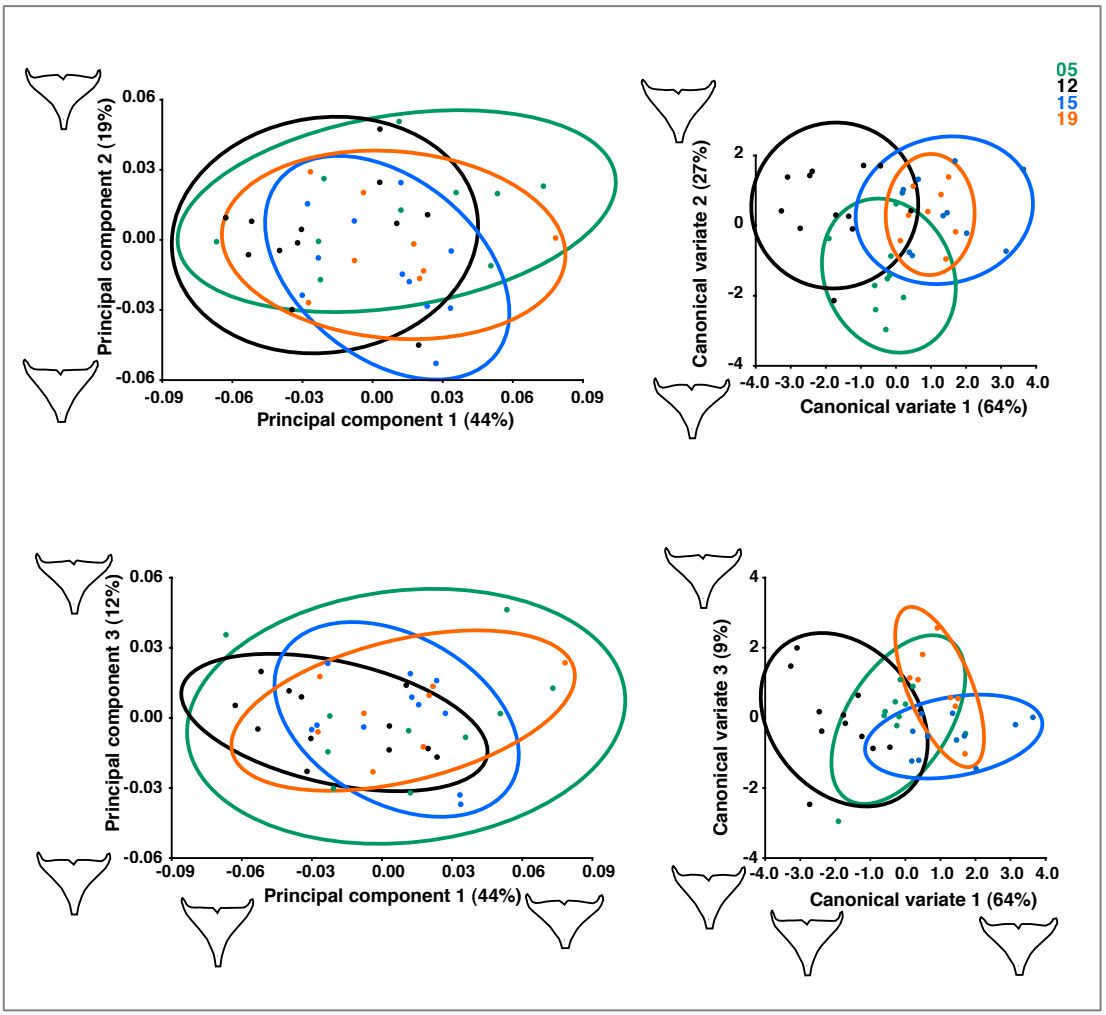


Figure 6.17. Shape analysis of lower pharyngeal jaw bone, *A. alcalica*.

Principal component analysis and canonical variate analysis for *A. alcalica* populations, as indicated by colours in top left.

Table 6.7. Pairwise distances of *A. alcalica* lower pharyngeal jaw bone shape.

Population	5	12	15	19
5 (n=10)	-	0.04	0.03	0.02
12 (n=12)	2.53*	-	0.04*	0.04*
15 (n=11)	2.49*	3.16*	-	0.02
19 (n=8)	2.30*	2.89*	8.00	-

Pairwise distances: Mahalanobis, below diagonal; Procrustes; above diagonal.

*P<0.05 (10,000 permutations)

Body size variation by population

While size effects were taken into account for the geometric morphometric analysis (i.e., a Procrustes fit was performed to remove scaling differences, and Procrustes coordinates were regressed against centroid size to remove any allometry of body shape with size), it was apparent that body size varied with habitat (Figure 6.18).

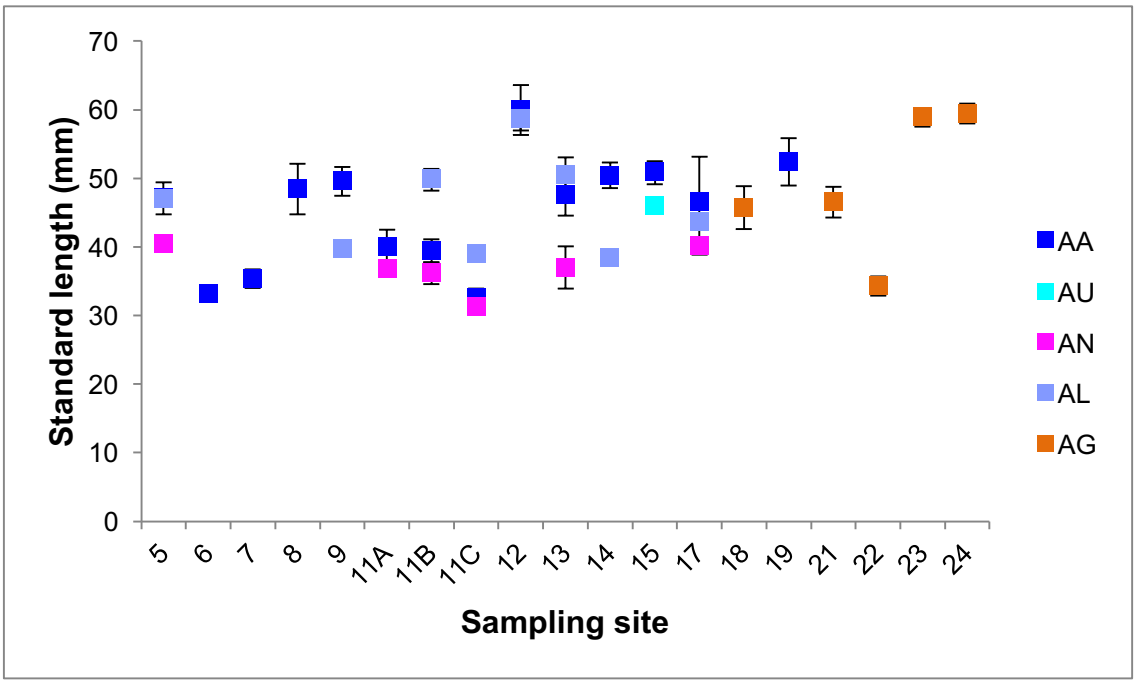


Figure 6.18. Standard length of all specimens included in the full geometric morphometric analysis (n=737), values are mean ± SEM for each species at each sampling site.

Lake Natron species showed the same pattern across populations, in that *A. alcalica* and *A. latilabris* were substantially larger than *A. ndalalani* in all within-site comparisons, although the largest *A. ndalalani* (found at sites 5, 13, 17) were larger than the smallest *A. alcalica* specimens (found at sites 6, 7and 11C). Standard length for *A. grahami* also varied across sites, with specimens from the introduced populations (sites 23 and 24, Lakes Elementeita and Nakuru) substantially larger than those from Lake Magadi. Despite these findings, body size did not appear correlated with habitat depth within-species for the limited number of sampling sites for which water depth was available (Figure 6.19).

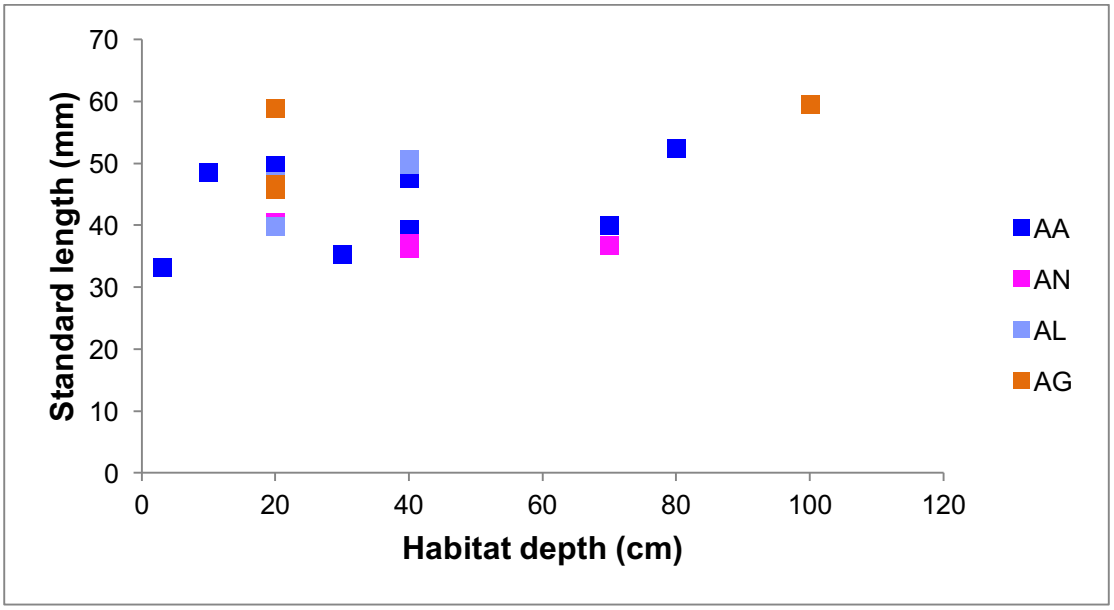


Figure 6.19. Body size (SL) plotted against habitat depth. No significant correlation was found for within-species comparisons.

Discussion

Although morphological variation between populations has previously been described in *Alcolapia* species (Seegers et al. 1999; Wilson et al. 2004), it has not been investigated using overall body shape (GMM) or within an ecological context. Here, morphometrics are used to define and quantify morphological variation across intraspecific populations, and paired with SIA to examine ecological variation.

Ecological differentiation among populations

In the present analysis, it is not possible to test whether *A. alcalica* and *A. grahami* exploit the same ecological niche and utilise similar trophic resources, as the species do not occur in sympatry and direct comparison of absolute values is not meaningful across sites. However the analysis does reveal that *A. grahami* inhabits a narrower niche based on carbon and nitrogen isotopes (Figure 6A.1; Table 6.2), suggesting greater dietary specialisation or less variation in basal source values. We may expect species occurring in monospecific populations to exhibit a greater niche breadth, with niche expansion via ecological release and driven by intraspecific competition, which has been demonstrated in three-spined stickleback (Bolnick *et al.* 2010). However, this is not observed in *A. alcalica* populations, where allopatric populations (e.g., at sites 6, 7, 8) do not exhibit a broader niche than in the sympatric populations (e.g., 5, 11, 12). The northern Natron population (site 19) exhibits the broadest niche of all populations, but as this site was a spring at the edge of the southern Ewaso Ngiro river, it seems likely that the breadth is due to environmental conditions and the input of nutrients from external sources, rather than the lack of interspecific competition. Broader niche space is also seen in *A. alcalica* populations that occur with only *A. latilabris* (sites 9 and 13), which may suggest that interspecific competition is weaker when only two species are in sympatry, or that *A. alcalica* does not directly compete with *A. latilabris* given its more specialised benthic foraging mode. However, overlap in niche space between *A. alcalica* and *A. latilabris* at site 14 could indicate incomplete segregation of niche exploitation in this population, and is mirrored by overlapping morphology between species at this site (Figure 6.14), which is discussed further below. For both *A. latilabris* and *A. ndalalani*, the broadest niche exploitation in each species is seen at site 17. This site had the lowest abundance of *A. alcalica* of all populations (only 4 *A. alcalica* specimens collected, 2 of which were suspected to be of hybrid origin),

so it may be that *A. latilabris* and *A. ndalalani* are experiencing character release in this population due to the lower abundance of the more generalist *A. alcalica*.

The narrow niche space of *A. grahami* in Lake Magadi may be due to the reduced resources available, with decreased hydrological input into the system (no perennial rivers or streams) relative to Natron, and more extreme conditions of Magadi springs (higher salinity levels and lower flow; Table 6A.1) restricting ecological niche available. From the stomach content analysis, it also seems that resource use as well as niche breadth is constrained between populations (Figure 6.6). It is surprising that the species has not undergone niche expansion in the translocated populations in Lakes Nakuru and Elementeita. These larger lakes offer substantially larger area and greater niche space, but it seems that *A. grahami* exhibits niche conservatism, which suggests lack of intraspecific competition driving expansion. This may be due to either insufficient time for the populations to grow large enough for substantial intraspecific competition (there being no other fish species in these lakes to create interspecific competition), or due to high predation pressure from water birds such as pelicans at these sites. Although specimens collected from these two lakes were larger than Lake Magadi specimens (Figure 6.18), it was not possible to collect any of the much larger fish thought to occupy the centre of the lake and grow up to 200mm (Vareschi 1979). Such specimens from Lake Nakuru are present in the collections of the Nairobi Museum, Kenya, and exhibit a deeper body shape (pers. obs.), so it may be that the species is undergoing divergence between shallow and deep(er) water morphs, with the shallow-water inhabitants conserving a niche space similar to that occupied in Lake Magadi.

Large variation in isotopic values is identified between all sampling sites, indicating the effect of different nutrient inputs across a relatively small environmental area and in similar habitats of the soda lake springs (Figures 6.3, 6.4 and 6A.1). For example, for both *A. alcalica* and *A. latilabris*, some of the greatest differences in SIA values were between sites 5 and 9 and 5 and 12, but these sites are some of the closest together (14 and 9km apart respectively). Although the soda lake sampling sites were depauperate in invertebrate and planktonic life, for the sites where baseline samples were available, isotopic values exhibited substantially different isotopic carbon signatures than the fish samples (Figure 6.5). This may suggest an additional invertebrate food source that was not collected during fieldwork. However, terrestrial-derived baseline samples, such as insects or plant matter, which appear to be consumed by at least *A. alcalica* (Figure 6.6), were not

sampled in the present study. Alternatively, the lack of corresponding baseline could indicate that fish are feeding on material of methanogenic origin, as methane-based sources are depleted in ^{13}C relative to other basal resources (Grey *et al.* 2004; Grey & Deines 2005; Harrod & Grey 2006), which may be prevalent within the system owing to the hypoxic conditions. Certainly, there are methanogenic bacteria present within the soda-lake basin (Grant 2006; Surakasi *et al.* 2007). Further analysis with more complete sampling of primary producers and invertebrates would be required to test this hypothesis.

Resource use exhibited substantial overlap between populations of *A. latilabris*, *A. ndalalani* and *A. grahami* (Figure 6.6, 6.7), while *A. alcalica* populations were differentiated. Although this may be in part due to the greater number of sites sampled relative to other species, it was apparent that northern and southern *A. alcalica* populations were feeding on different resources (Figure 6.6; 6.7). While sites 5, 11, 12 (southern, sympatric with other species) exhibited high proportion of cellulose consumed (Figure 6.6), the northern monospecific sites 15 and 19 had very little cellulose contribution to diet, suggesting that in sites where *A. alcalica* occurs alone without the benthic specialists *A. latilabris* and *A. ndalalani*, it occupies a different niche space, feeding mainly on algae and cyanobacteria. Site 19 individuals had a high proportion of grit/sand, also suggesting a benthic feeding strategy, although this site was deep and slow-flowing, so there may also be environmental factors contributing to this. Site 19 was a spring at the edge of the Ewaso Ngiro, and substantially deeper and cooler than the other collection samples (Table 6A.1). Thus the differentiation of this site in genomic and ecomorphological analysis (chapter three, and the present chapter) from other populations could be a result of environmental variables as well as the geographical separation from other populations. A comparison could be made with the other major inflowing river in Lake Natron, River Peninj, on the Western shore, although it was not visited in the present study, and to my knowledge has not been surveyed for the presence of *Alcolapia* populations. The separation of northern/southern sites is also supported by the differentiated morphology of pharyngeal jaws (Figure 16.17, Table 6.7), although this was only based on a small sample number.

The *A. alcalica* upturned-mouth morph at site 15 exhibited a small proportion by diet of cellulose and fish scales (Figure 6.6), which was not seen in the terminal-mouth morph, but the two morphs mainly fed on algae and substantially overlapped in diet (Figure 6.7). The overlap between these morphs in niche space (based both on SIA values and stomach contents) was surprising given the presumed separation

by ecological trait of mouth position. This result is likely explained by the sex ratio results, discussed below.

Intraspecific morphological variation

Intraspecific variation appears greatest within *A. grahami*, and while this is likely driven by the inclusion of the introduced populations, morphological differentiation at the population level has previously been reported in Lake Magadi *A. grahami* between different lagoons (Wilson et al. 2004). Re-running the CVA without the introduced populations produced the same results in pairwise comparisons of Lake Magadi populations, with no differentiation between sites 18 and 21, but significant differentiation between site 22 and the other two populations. This reflects the geographic separation of these three sites. Although it should be noted that running CVA on small numbers of groups is more likely to produce differentiated groups (Strauss 2010), so is probably not comparable to the other species comparisons containing larger numbers of populations.

The sub-analyses of populations that appeared divergent in within-species PCA revealed a close relationship of *A. alcalica* and *A. latilabris* at site 14, which overlapped both in ecological niche based on SIA (Appendix Figure 6A.2) and in morphospace (Figure 6.14). The individuals from this site were darker in colouration than at other sites and while *A. latilabris* individuals had thick lips and wide mouths these did not appear as distinctly down-turned as in other populations. The overlapping morphology and ecology suggests that *A. alcalica* and *A. latilabris* are incompletely segregated in niche space at this site, and could be indicative of incomplete reproductive isolation. Unfortunately no genetic data was available from this population in order to test for possible hybridisation as a cause for intermediate phenotype. The site is approximately 600m from site 15 of (Seegers & Tichy 1999), where an 'intermediate form of *A. alcalica*/*A. latilabris*' was recorded, so it seems likely that this is another site with extensive hybridisation (see discussion of site 17 in chapters three and five).

The difference in sex ratios between populations was most marked at site 15 (Figure 6.8). This observation suggested that the terminal and upturned-mouth morphs at this site in fact represent sexual dimorphism, rather than ecological morphotypes as had previously been suspected. That the 'morphs' overlap in stable isotope ratio (chapter five) and stomach contents (Figures 6.6, 6.7) indicates that the morphological differentiation does not result in ecological differentiation. Further

analysis of morphology by sex, on body shape (Figure 6.15) and cranial landmarks alone (Figure 6.16) indicated that sexes are differentiated at site 15 but not for other populations of *A. alcalica* or other species. Assuming that this is a case of sexual dimorphism of the same species at this site (and not biased sampling), it is striking that it is not seen in other populations. Although environmental readings were not available from site 15, field observations and photographs show a marked difference from other sites, where site 15 was a stagnant, deep, ditch, compared to the shallow, rapid springs seen at other sites. Furthermore, site 15 was substantially further away from the Lake shoreline (2.1km) than most other sites (e.g., site 5: 490m; site 11: 550m) and of all the populations, it was the only site which did not flow directly into a lagoon or area of open water. This suggests that the population at site 15 is likely to be more isolated from other populations. The locality is on the site of disused magnesite mine, and with a dirt road leading to Magadi town. The settlement is no longer inhabited, but the presence of a previous settlement means that the possibility that the *Alcolapia* population was introduced here (rather than naturally colonising from other populations) cannot be ruled out. However, given the phylogenetic results (chapter three), it is apparent that the population is not *A. grahami* introduced from Magadi (the closest town). While cranial shape differences between sexes have been reported in cichlids in Lake Malawi and thought to be an adaptation to mouthbrooding (Herler *et al.* 2010), these were represented by differences in buccal cavity shape, while mouth position variation was seen between ecologically divergent populations (Herler *et al.* 2010). Increased mouth angle has been attributed to sexual dimorphism in *Astatotilapia burtoni* from Lake Tanganyika; but male lake morphs of this species also exhibited a superior mouth relative to stream populations (Theis *et al.* 2014). These observations suggest a divergence of lentic-lotic conditions, which is pertinent to site 15 in this study given the lack of flow at this site. A superior mouth position is reported in *A. grahami* from Little Magadi compared to Lake Magadi populations (Wilson et al. 2004), although it is not known if this is correlated with environmental variations. Craniofacial differentiation between sexes has been suggested to represent 'nested variation' by ecological sexual dimorphism in a recent analysis of a Lake Malawi mouthbrooder (Parsons *et al.* 2015). The authors suggested that the shallower head profile of female *Labeotropheus fuelleborni* could be explained by more time spent feeding in the water column, while aggressive males were segregated to defending territorial sites on the substrate (Parsons *et al.* 2015).

Conclusions

An integrated dataset to investigate morphological and ecological differentiation within species provides a preliminary examination of population divergence within the *Alcolapia* group. The striking results of differentiation within *A. alcalica* certainly warrant further investigation. Increased sampling at site 15 would be able to confirm the proposal of ecological sexual divergence, and field observations could confirm whether sexes are segregated within the water column. Additionally, environmental measurements would clarify the environmental influence on the cranial morphological differentiation relative to other sites. The morphological overlap at site 14 warrants genetic investigation to determine the extent of hybridisation at this site. Another site identified by Seegers *et al.* (2001) (not visited in the present study, but south of site 14 in this study) was inhabited by individuals with intermediate morphology of *A. alcalica* and *A. ndalalani*, suggesting that populations found on the south-east coast experience hybridisation

Future research may focus on the following areas of interest:

- Further sampling and genomic investigation of the extent of hybridisation in populations on the southeast shore (sites 14 and 17 from the present study, additional site from Seegers et al. 2001)
- Adaptive divergence of populations in areas of inflowing rivers and streams: sampling at Peninj river on the West shore
- Further investigation of sex ratio and sexual dimorphism at site 15 of the present study
- Increased SIA baseline sampling to characterise the food web and investigate the hypothesis of methanogenesis being an important contributor to the food chain
- Investigation of sex ratio in natural vs. introduced populations of *A. grahami*
- Examination of the adaptive landscape (by testing relative individual fitness) across environmental gradients, for example between cooler/deeper water (site 19) and shallower/warmer sites, and between lentic/lotic conditions

References

Abucay JS, Mair GC, Skibinski DOF, Beardmore JA (1999) Environmental sex determination: the effect of temperature and salinity on sex ratio in *Oreochromis niloticus* L. *Aquaculture*, **173**, 219–234.

Barluenga M, Stolting KN, Salzburger W *et al*. (2006) Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*, **439**, 719–723.

Bezault E, Feder C, Derivaz M, *et al*. (2007) Sex determination and temperature-induced sex differentiation in three natural populations of Nile tilapia (*Oreochromis niloticus*) adapted to extreme temperature conditions. *Aquaculture*, **272**, S3–S16.

Binning SA, Chapman LJ (2010) Is intraspecific variation in diet and morphology related to environmental gradients? Exploring Liem's paradox in a cichlid fish. *Integrative Zoology*, **5**, 241–255.

Bolnick DI, Ingram T, Stutz WE *et al.* (2010) Ecological release from interspecific competition leads to decoupled changes in population and individual niche width. *Proceedings of The Royal Society B: Biological Sciences*, **277**, 1789–1797.

Caro LM, Caycedo-Rosales PC, Bowie RCK *et al*. (2013) Ecological speciation along an elevational gradient in a tropical passerine bird? *Journal of Evolutionary Biology*, **26**, 357–374.

Cutwa MM, Turingan RG (2000) Intralocality variation in feeding biomechanis and prey use in *Archosargus probatocephalus* (Teleostei, Sparidae), with implications for the ecomorphology of fishes. *Environmental Biology of Fishes*, **59**, 191–198.

Erlandsson A, Ribbink AJ (1997) Patterns of sexual size dimorphism in African cichlid fishes. *South African Journal of Science*, **93**, 498–508.

Fitzpatrick SW, Gerberich JC, Kronenberger JA, Angeloni LM, Funk WC (2015) Locally adapted traits maintained in the face of high gene flow. *Ecology Letters*, **18**, 37–47.

Ford AGP, Dasmahapatra KK, Rüber L *et al.* (2015) High levels of interspecific gene flow in an endemic cichlid fish adaptive radiation from an extreme lake environment. *Molecular Ecology*, doi: 10.1111–mec.13247.

Futuyma DJ, Moreno G (1988) The evolution of ecological specialization. *Annual Review of Ecology and Systematics*, **19**, 207–233.

Grant WD (2006) Alkaline environments and biodiversity. In: (Eds Gerday C, Glansdorff N), Extremophiles. pp. 1–20. UNESCO, Oxford.

Grey J, Deines P (2005) Differential assimilation of methanotrophic and chemoautotrophic bacteria by lake chironomid larvae. *Aquatic Microbial Ecology*, **40**, 61–66.

Grey J, Kelly A, Jones RI (2004) High intraspecific variability in carbon and nitrogen stable isotope ratios of lake chironomid larvae. *Limnology and Oceanography*, **49**, 239–244.

Harrod C, Grey J (2006) Isotopic variation complicates analysis of trophic relations within the fish community of Plußsee: a small, deep, stratifying lake. *Archiv für Hydrobiologie*, **167**, 281–299.

Hendry AP, Taylor EB, McPhail JD (2002) Adaptive divergence and the balance between selection and gene flow: lake and stream stickleback in the Misty system. *Evolution*, **56**, 1199–1216.

Herler J, Kerschbaumer M, Mitteroecker P *et al*. (2010) Sexual dimorphism and population divergence in the Lake Tanganyika cichlid fish genus *Tropheus*. *Frontiers in Zoology*, **7**, 4.

Hickley P, Muchiri M, Britton R, Boar R (2008) Economic gain versus ecological damage from the introduction of non-native freshwater fish: Case studies from Kenya. *The Open Fish Science Journal*, **1**, 36–46.

Hulsey CD, Roberts RJ, Loh YHE *et al*. (2013) Lake Malawi cichlid evolution along a benthic/limnetic axis. *Ecology and Evolution*, **3**, 2262–2272.

Jones FC, Chan YF, Schmutz J *et al.* (2012) A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Current biology*, **22**, 83–90.

Kleindorfer S, Winkler H, Sulloway FJ (2006) Adaptive divergence in contiguous populations of Darwin's small ground finch (*Geospiza fuliginosa*). *Evolutionary Ecology Research*, **8**, 357–372.

Langin KM, Sillett TS, Funk WC *et al.* (2015) Islands within an island: Repeated adaptive divergence in a single population. *Evolution*, **69**, 653–665.

Liem KF (1980) Adaptive significance of intra- and interspecific differences in the feeding repertoires of cichlid fishes. *American Zoologist*, **20**, 295–314.

Maan ME, Sefc KM (2013) Colour variation in cichlid fish: developmental mechanisms, selective pressures and evolutionary consequences. *Seminars in Cell and Developmental Biology*, **24**, 516–528.

Maithya J, Njiru M, Okeyo-Owuor JB, Gichuki J (2012) Some aspects of the biology and life-history strategies of *Oreochromis variabilis* (Boulenger 1906) in the Lake Victoria Basin. *Lakes & Reservoirs: Research & Management*, **17**, 65–72.

Moser D, Roesti M, Berner D (2012) Repeated lake-stream divergence in stickleback life history within a central european lake basin. *PLoS ONE*, **7**, e50620.

Nichols P, Genner MJ, van Oosterhout C *et al.* (2014) Secondary contact seeds phenotypic novelty in cichlid fishes. *Proceedings of The Royal Society B: Biological Sciences*, **282**, 2272.

Nosil P (2012) *Ecological speciation*. Oxford University Press, Oxford.

Oliveira RF, Almada VC (1995) Sexual dimorphism and allometry of external morphology in *Oreocrhomis mossambicus*. *Journal of Fish Biology*, **46**, 1055–1064.

Papah MB, Kisia SM, Ojoo RO *et al.* (2013) Morphological evaluation of spermatogenesis in Lake Magadi tilapia (*Alcolapia grahami*): A fish living on the edge. *Tissue and Cell*, **45**, 371–382

Parsons KJ, Wang J, Anderson G, Albertson RC (2015) Nested levels of adaptive divergence: The genetic basis of craniofacial divergence and ecological sexual dimorphism. *G3: Genes|Genomes|Genetics*, doi: 10.1534–g3.115.018226.

Roberts N, Taieb M, Barker P *et al.* (1993) Timing of the Younger Dryas event in East Africa from lake level changes. *Nature*, **366**, 146–148.

Schluter D (1993) Adaptive radiation in sticklebacks: Size, shape and habitat use efficiency. *Ecology*, **74**, 699–709.

Schluter D (2009) Evidence for ecological speciation and Its alternative. *Science*, **323**, 737–741.

Schütz D, Taborsky M (2000) Giant males or dwarf females: what determines the extreme sexual size dimorphism in *Lamprologus callipterus*? *Journal of Fish Biology*, **57**, 1254–1265.

Seegers L, Tichy H (1999) The *Oreochromis alcalicus* flock (Teleostei: Cichlidae) from Lake Natron and Magadi, Tanzania and Kenya, with description of two new species. *Ichthyological Explorations of Freshwaters*, **10**, 97–146.

Seehausen O, Terai Y, Magalhaes IS *et al.* (2008) Speciation through sensory drive in cichlid fish. *Nature*, **455**, 620–626.

Strauss RE (2010) Discriminating groups of organisms. In: (Ed. Elewa AMT) Morphometrics for Nonmorphometricians. Springer-Verlag, Berlin. pp. 73–91.

Surakasi VP, Wani AA, Shouche YS, Ranade DR (2007) Phylogenetic analysis of methanogenic enrichment cultures obtained from Lonar Lake in India: Isolation of *Methanocalculus* sp. and *Methanoculleus* sp. *Microbial Ecology*, **54**, 697–704.

Theis A, Ronco F, Indermaur A, Salzburger W, Egger B (2014) Adaptive divergence between lake and stream populations of an East African cichlid fish. *Molecular Ecology*, **23**, 5304–5322.

Vareschi E (1979) The ecology of Lake Nakuru (Kenya). II. Biomass and spatial distribution of fish (*Tilapia grahami* Boulenger = *Sarotherodon alcalicum grahami* Boulenger). *Oecologia*, **37**, 321–335.

Wilson PJ, Wood CM, Walsh PJ *et al.* (2004) Discordance between genetic structure and morphological, ecological, and physiological adaptation in Lake Magadi tilapia. *Physiological and Biochemical Zoology*, **77**, 537–555.

Zaccara S, Crosa G, Vanetti I *et al.* (2014) Genetic and morphological analyses indicate high population mixing in the endangered cichlid *Alcolapia* flock of East Africa. *Conservation Genetics*, **15**, 429–440.

Østbye K, NAESJE TF, Bernatchez L, Sandlund OT, HINDAR K (2005) Morphological divergence and origin of sympatric populations of European whitefish (*Coregonus lavaretus* L.) in Lake Femund, Norway. *Journal of Evolutionary Biology*, **18**, 683–702.

Appendix 6

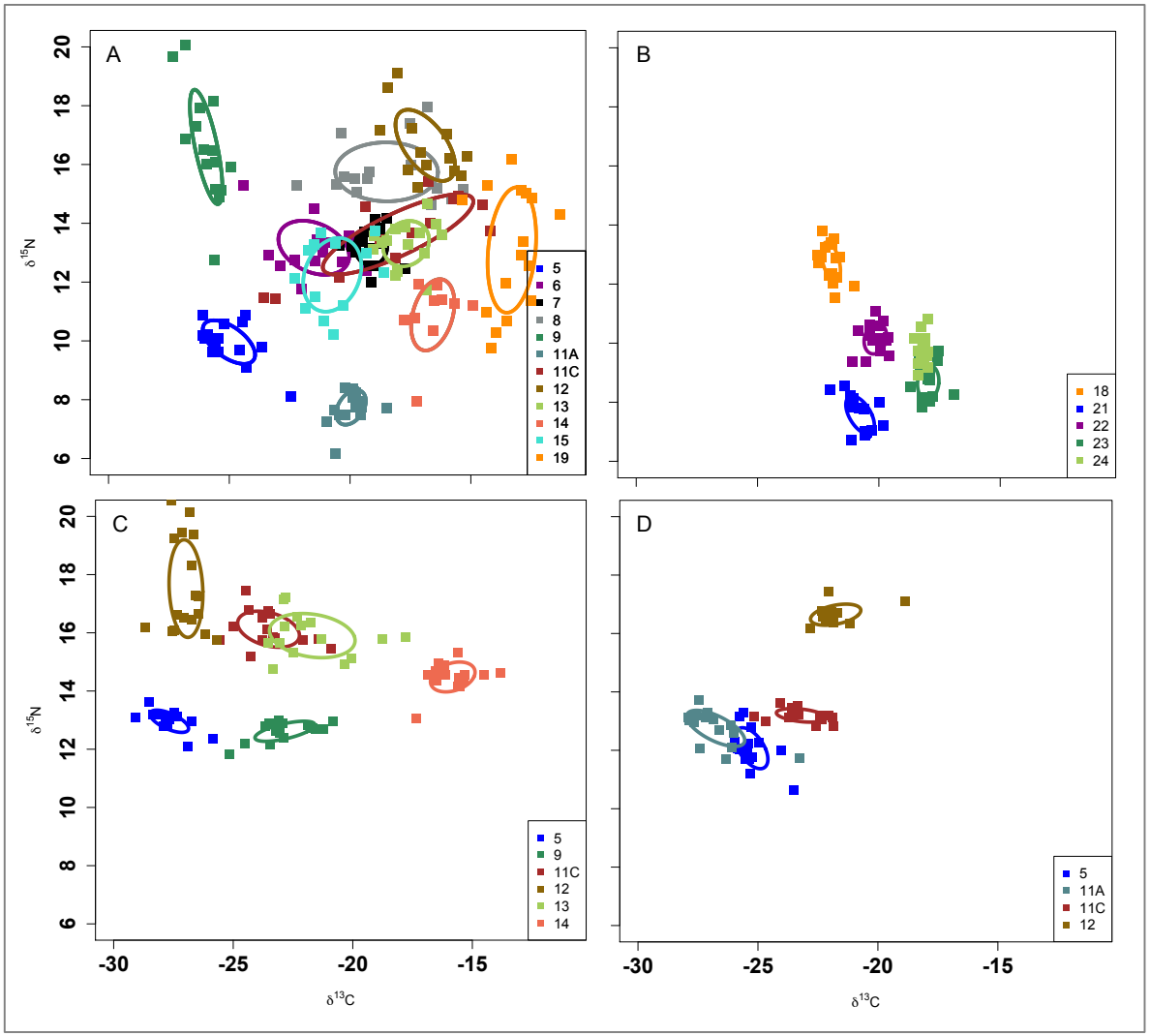


Figure 6A.1. Stable isotope values plotted per species.

A) *A. alcalica*; B) *A. grahami*; C) *A. latilabris*; D) *A. ndalalani*

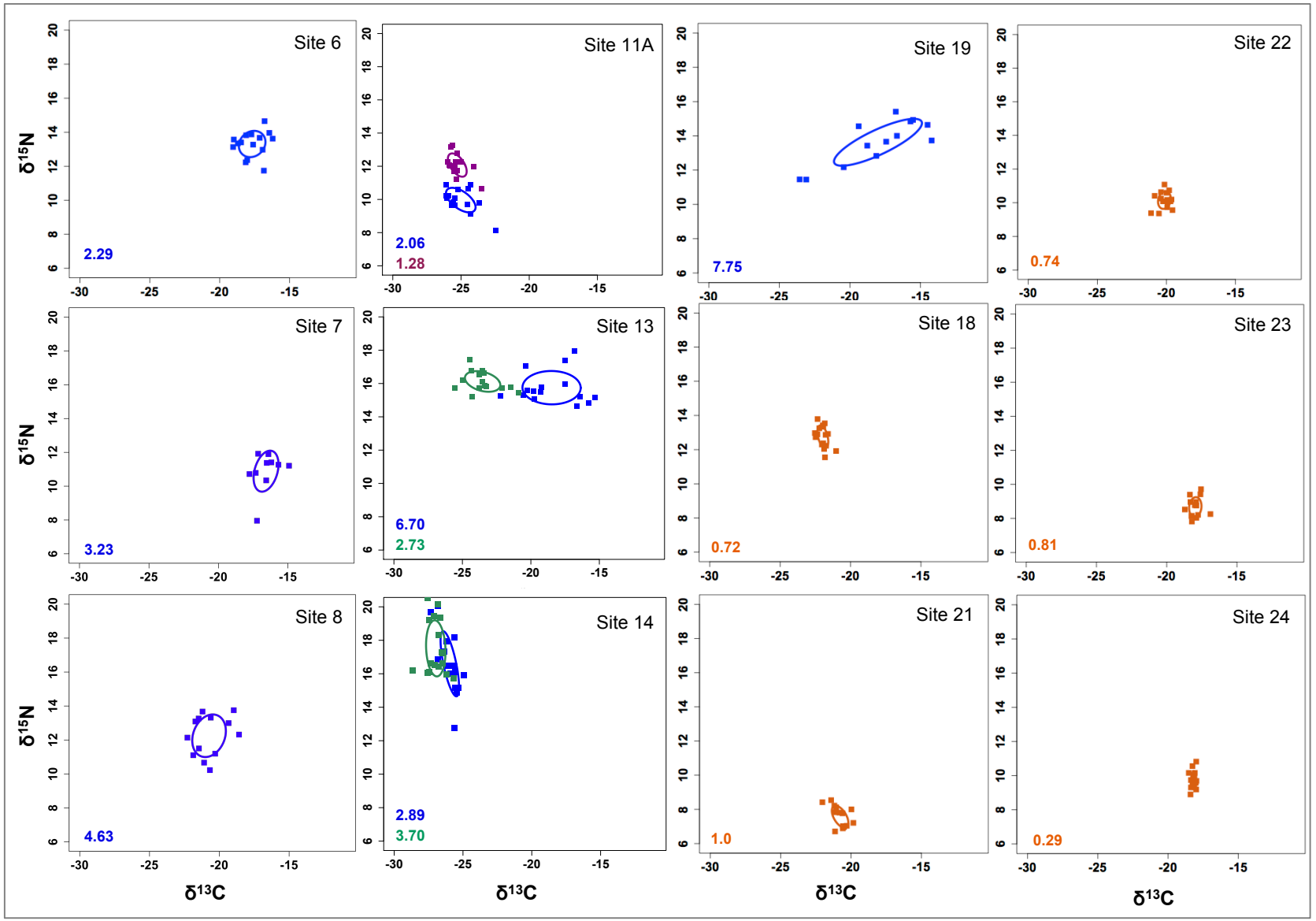


Figure 6A.2. Stable isotope plots for additional sites not included in chapter five.

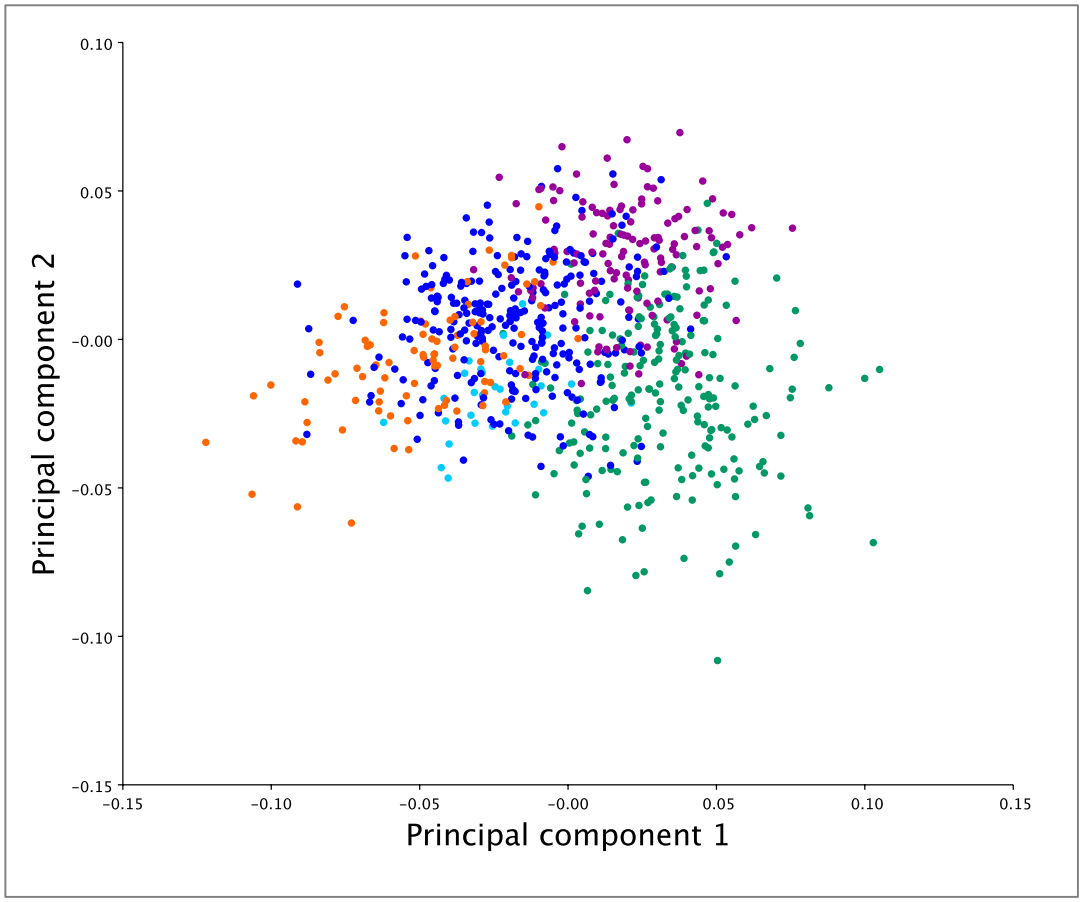


Figure 6A.3. PCA of full dataset (n=734) coloured by species.

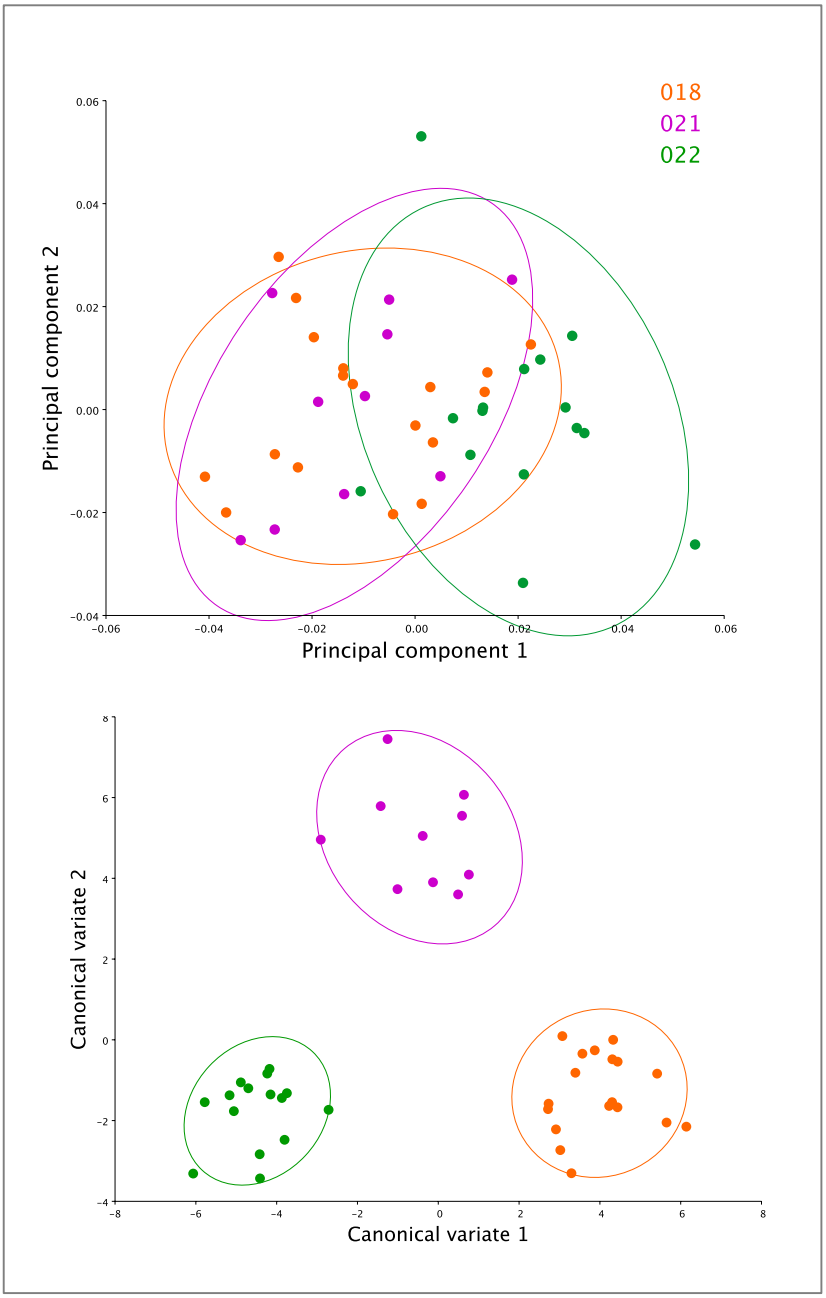


Figure 6A.4. PCA and CVA of *A. grahami* including Lake Magadi populations only.

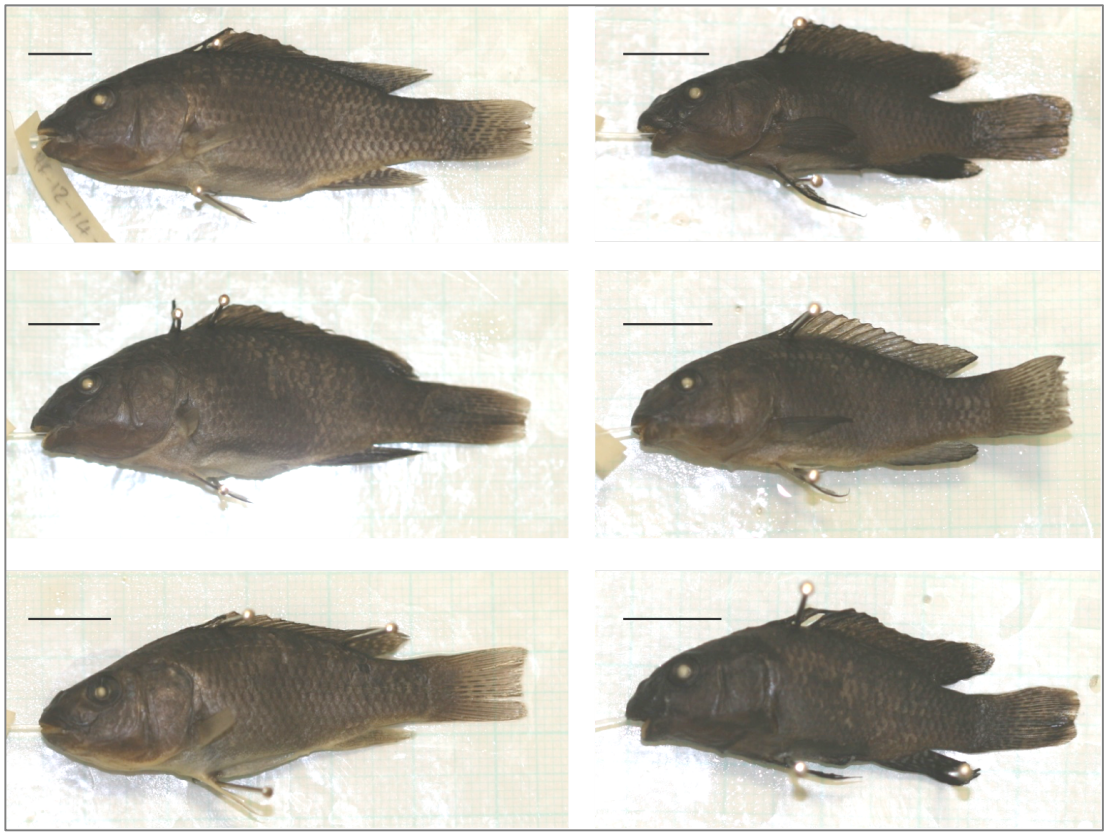


Figure 6A.5. Photographs of individuals from site 14.

Photographs of individuals designated *A. alcalica* and *A. latilabris* at site 14, showing lack of distinctive species characteristics seen at other sites. Initial species labelling relied on width of mouth rather than typical species traits. Left-hand column contains individuals designated as A. alcalica; right-hand column: *A. latilabris*.

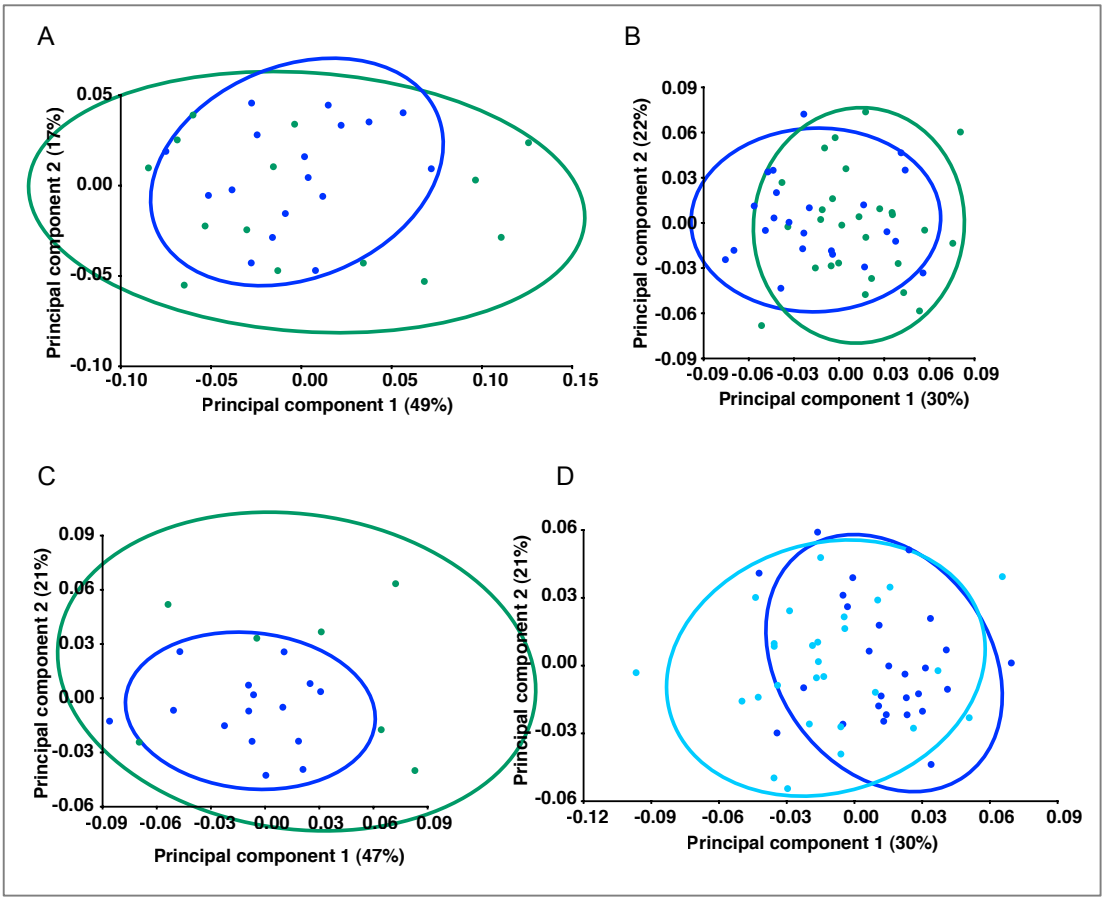


Figure 6A.6. PCA of cranial landmarks by population.

Blue, male; Green, female; light blue, upturned-mouth morph (female). A) *A. latilabris* (sites 5 and 12); B) *A. ndalalani* (sites 5 and 12); C) *A. alcalica* (sites 5 and 12); D) *A. alcalica* (site 15) terminal and upturned mouth morphs.

Table 6A.1 Environmental readings for sampling sites.

GPS coordinates for each site are given in Table 6.1.

Site	Location	Habitat	Depth (cm)	°C	pH	Con. (ppt)	DO (mg/L)	Flow (RPM)
005	South Natron	Spring	15–20	29.8	9.83	2.6	1.8	122
006	West Natron	Spring	3	28.4	9.96	>10	1.9	
007	West Natron	Spring	30	37.3	10.08	4.58	0.2	73
008	West Natron	Spring	10	32.2	8.81	0.7	6.5	522
009	West Natron	Spring	20	33.6	9.97	1.8	6.1	498
11A	SE Natron	Lagoon	40	33.1	9.83	>10	0.08	446
11B	SE Natron	Rapids	70	33.6	8.77	3.4	6.6	1165
11C	SE Natron	Spring	-	-	-	-	-	-
012	South Natron	Spring	-	-	-	-	-	-
013	South Natron	Spring	10–40	28	9.7	2.4	2.2	178
014	SE Natron	Spring	-	-	-	-	-	-
015	East Natron	Spring	-	-	-	-	-	-
017	East Natron	Spring	-	-	-	-	-	-
018	South Magadi	Hot springs	20	39.8	9.53	>10	11.9	58
019	North Natron	River edge	50	24.5	10.1	2.21	5.8	16
021	NW Magadi	Lake edge	5–25	41.1	9.41	>10	16.6	-
022	East Magadi	Springs	10	30	9.92	>10	0.5	-
023	Elementeita	Lake, flooded	20–30	29.9	9.73	3	5.2	-
024	Nakuru	Lake, flooded	5–105	32.5	10.1	4.7	21.9	-

Cond.: conductivity. Sites 11 A, B, C were points of increasing elevation along the same spring. Inter-site distances were: 11A-11B: 300m; 11B-11C: 500m.

Table 6A.2. Eigenvalues for PCA on the whole dataset (n=734)

PC	Eigenvalues	% Variance	% Cumulative
1	0.00136182	38.10	38.10
2	0.00075121	21.02	59.12
3	0.00021357	5.98	65.10
4	0.00017717	4.96	70.05
5	0.00015867	4.44	74.49
6	0.0001458	4.08	78.57
7	0.00012625	3.53	82.11
8	0.00011232	3.14	85.25
9	0.00008723	2.44	87.69
10	0.00007052	1.97	89.66
11	0.00005809	1.63	91.29
12	0.00004473	1.25	92.54
13	0.0000417	1.17	93.71
14	0.0000363	1.02	94.72
15	0.00003263	0.91	95.63
16	0.00002797	0.78	96.42
17	0.00002105	0.59	97.01
18	0.00001764	0.49	97.50
19	0.00001575	0.44	97.94
20	0.00001314	0.37	98.31
21	0.00001176	0.33	98.64
22	0.00001095	0.31	98.94
23	0.00000978	0.27	99.22
24	0.00000868	0.24	99.46
25	0.00000711	0.20	99.66
26	0.00000564	0.16	99.82
27	0.00000535	0.15	99.97
28	0.00000121	0.03	100

Chapter seven

Conclusions and future directions

Overview of findings

Despite clear morphological differences and unique physiological adaptations in soda lake cichlids, previous genetic work has been unable to resolve relationships within the *Alcolapia* radiation. In chapter three, an extensive genomic dataset containing dense sampling of the entire *Alcolapia* was presented and the phylogeny of the entire radiation addressed. The use of densely sampled genome-wide SNPs achieved resolution in three constituent species (excluding one anomalous site of suspected hybridisation). The phylogeny demonstrated that *A. alcalica* from southern and northern Natron lagoons form separate clades. Population genomics approaches demonstrated high levels of gene flow between species in populations bordering the southern lagoon of Lake Natron. Interspecific differentiation was low (F_{ST}=0.04-0.20), but comparable to levels of differentiation seen in other young fish radiations. In contrast to the close relationships of *Alcolapia* species, considerably more differentiation was found within *O. amphimelas* between the geographically separated Lake Eyasi and Manyara populations than within *Alcolapia,* highlighting the possibility of cryptic diversity within the other soda lakes of East Africa.

The results of chapter three guided the analysis in the remaining chapters (for example, in the delineation of northern and southern *A. alcalica* clades) and the phylogeny was employed in chapter five to consider the correlation of ecomorphological relationships with genomic patterns. Evidence for narrow regions of genomic differentiation containing outlier loci was presented in chapter four, for all *Alcolapia* species comparisons and intraspecific population comparisons of *A. alcalica*. Other intraspecific comparisons exhibited peaks of elevated F_{ST} in genomic scans, but no clear detection of outlier loci.

Building on the results of the genomic analysis, chapter five integrated several datasets to examine ecomorphological differentiation between species. Morphology has featured prominently in the species descriptions and taxonomic treatment of this genus (e.g., Thys van den Audenaerde 1968; Trewavas 1983; Seegers & Tichy 1999) as well as in more recent population level analyses (Zaccara *et al.* 2014), but the ecology of these species has not been previously addressed. Indication of diet

was previously restricted to field observations and stomach contents of *A. grahami* (Coe 1966). Here, stable isotope and analysis and stomach contents analysis was employed to provide comparative data for all described species. All species were differentiated by trophic niche, although *A. latilabris* and *A. ndalalani* displayed only fine-scale differentiation based on foraging mode in benthic feeding. By contrast, the species divergence seen in morphological body shape patterns (particularly trophic morphology) was large, while differences in pharyngeal jaw shape were not significant, reflecting the fine scale distinction of dietary partitioning. Substantial differences in ecomorphology were not reflected in genomic differentiation. SURFACE analysis suggested that the similarity in body shape between *A. alcalica* and *A. grahami* was an ancestral rather than convergent form. The analysis also revealed that there were three clear regime shifts within the radiation, in particular in the diversification of the sub-terminal mouth morphologies of *A. latilabris* and *A. ndalalani*.

Ecomorphology was further explored in chapter six at an intraspecific level. Trophic niche breadth (from stable isotope analysis) indicated that *A. alcalica* exploited the broadest niche across its range, and coupled with stomach contents analysis indicated that populations occurring in sympatry with other species exploit additional resources (cellulose) than those occurring in allopatry. *Alcolapia grahami* exhibited a conserved niche, exploiting the same resources in varying environmental conditions, suggesting that the lack of interspecific competition has resulted in a lack of niche expansion. The additional inclusion of sex data to consider sex ratios revealed a proposed instance of ecological sexual divergence at site 15. Although only preliminary results of a small dataset, these results suggest that isolated populations separated from the main lagoons are experiencing more rapid adaptive divergence than other populations. Furthermore, the inclusion of additional sampling sites allowed the identification of a hybridisation site or additional morph in the intermediate *A. alcalica* / *A. latilabris* population at site 14.

Taken together, these results of low genomic differentiation, high gene flow, substantial morphological divergence and fine-scale ecological partitioning suggest a species flock of very recent origin undergoing adaptive divergence.

Alcolapia as an example of adaptive radiation and ecological speciation?

An overarching aim of this thesis was to investigate whether the *Alcolapia* species flock represents an adaptive radiation. Requirements to define species pairs as

examples of ecological speciation include i) phenotypic differentiation driven by natural selection and ii) reproductive isolation as a consequence of divergent natural selection (Faria *et al.* 2014). Although several studies present data consistent with ecological speciation, few have shown unambiguous cases (Hendry 2009; Faria *et al.* 2014). Here, the pattern of low genomic divergence with clear trophic and morphological differentiation is certainly suggestive of recent adaptive radiation with rapid ecological speciation, and indicative of filling of niche space within this system. However, while the present analysis is congruent with a hypothesis of ecological speciation, particularly with outlier analysis suggesting divergent selection between species, further work would be required to demonstrate that adaptive divergence reduces gene flow.

The defining criteria of adaptive radiation are slightly more specific: i) common ancestry, ii) a correlation between phenotype and environment, iii) traits leading to a fitness advantage in a particular environment (trait utility), and iv) rapid speciation (Schluter 2000). Common ancestry and rapid speciation are demonstrated in the phylogenomic analysis of chapter three. A phenotype-environment correlation is supported here by the trophic morphology correlation with diet (stable isotope) and inferred foraging mode (benthic/surface vegetation in stomach contents analysis, and particle size in benthic feeders). This is further supported by anecdotal field observations (pers. obs.) of surface and water column feeding of *A. alcalica*, benthic scraping by *A. latilabris* and benthic picking by *A. ndalalani*. Trait utility (fitness advantage) is not explicitly tested here, but the inference that *A. alcalica* is predominantly a benthic feeder in allopatric sites, but is displaced to a substantial proportion of surface feeding in sympatric sites suggests that the sub-terminal mouth morphology of *A. ndalalani* and *A. latilabris* confers a competitive advantage in benthic feeding. The identification of distinct regime shifts in chapter five also supports of the conclusion of rapidly speciated lineages filling divergent niche space.

Although further empirical work will be required to fully identify these patterns (see below for further research goals), these analyses lay the foundation for future studies to establish *Alcolapia* as a useful case study for ecological speciation and adaptive divergence.

Alcolapia species flock as a study system for speciation

The *Alcolapia* flock represents a young, small-scale radiation with several incipient species. Although the present study demonstrates that the *Alcolapia* lineages may not be strongly reproductively isolated (chapter three), with significant levels of

admixture between species, they may be considered species under the genotypic clusters species concept (Mallet 1995). Given the recent diversification of the species flock, there is likely to be some level of incomplete lineage sorting, and yet nearly all samples in the phylogenomic analysis sort by species first, and then sampling site (chapter three), suggesting there are existing segregating sites between species. It seems unlikely that these differences merely describe population differentiation (via drift or local adaptation) given that divergence is maintained in geographical contact (deemed a 'critical test' of incipient speciation, Seehausen & Wagner 2014). Furthermore, the occurrence of ongoing gene flow and admixture in founding populations is increasingly being found in adaptive radiations (e.g., Lamichhaney *et al.* 2015), and several fish radiations are thought to have emerged from a 'hybrid swarm' origin (Seehausen 2004; Hudson *et al.* 2011). As such, the *Alcolapia* species flock appears to be at a very early stage of speciation and offers an excellent system to investigate processes generating biodiversity. Incipient species may be the most useful for examining generation of reproductive isolation, where barriers that contributed to speciation (rather than arose after speciation was complete) can be tested (Coyne & Orr 2004). The *Alcolapia* radiation provides an excellent setting in which to conduct research on speciation processes, as it includes lineages and populations at differing levels of separation such that the processes involved at different stages of the speciation continuum can be examined. These comparisons include: allopatrically separated non-sister species (*A. grahami* and *A. latilabris*/*A. ndalalani*), sympatrically occurring non-sister species (*A. alcalica* and *A. latilabris*/*A. ndalalani*); sympatric sister species (*A. ndalalani* and *A. latilabris*); intraspecific allopatric populations of recent separation (*A. alcalica* north/south clades) and very recent separation (*A. grahami* Magadi/Nakuru lakes); and intraspecific sympatric morphs of colour and trophic morphology (*A. alcalica* yellow/blue colour morphs and terminal/upturned mouth morphs). The presence of several species with differing distributions allow several comparisons to be conducted, including sympatry/allopatry (*A. alcalica* in southern and northern lagoons, co-occurring with two, one, or no other species); the ecological effect of translocation to novel environments (*A. grahami* in Lakes Nakuru and Elementeita); and impact of population connectivity, comparing a gradient of highly connected populations (Lake Natron southern lagoon) towards more isolated populations (eastern Natron populations) and highly isolated populations (*A. alcalica* in north Natron). Investigation of ecological gradients, varying population densities and recurrent connection/isolation of populations is particularly pertinent given the recent

application of clinal theory to speciation and adaptive radiation research (Abbott *et al*. 2013; Seehausen 2015)

Finally, colonisation inference may be more straightforward in this young and geographically restricted system than in larger water bodies such as the African Great Lakes with older radiations and greater species diversity. The endemism of *Alcolapia* within the Natron/Magadi basin means that repeated colonisation scenarios or continuing introgression from external sources (e.g., Schliewen *et al.* 2006; Martin *et al*. 2015) are unlikely.

Conservation priority

The unique physiological adaptations to extreme environmental conditions in *Alcolapia*, their endemism and geographic exclusion from other *Oreochromis* species, as well as a fragile ecosystem sensitive to anthropogenic change, create a conservation priority for these fishes. All *Alcolapia* species are categorised as endangered or vulnerable on the IUCN red list (Bayona 2006; Bayona & Akinyi 2006) and populations are potentially threatened by planned development of an ash mining plant at Lake Natron with concurrent development of water extraction and infrastructure (Kadigi *et al.* 2012). The results of the present analysis identify additional morphs (site 14) and isolated populations (site 15) previously unreported, and suggest that further surveying of the area is required to highlight areas of most concern with regards to environmental change or habitat degradation.

Future goals

The present thesis provides a preliminary characterisation of the *Alcolapia* species flock, and identifies several areas that warrant further investigation. The phylogenomic analysis resolved three constituent species, but lacked clarity in the placement of the geographically widespread *A. alcalica* clades. Additional genomic resources of more densely sampled SNPs, such as whole genome resequencing may provide better resolution of these relationships. A whole genome sequence dataset has been sequenced and is currently in analysis to consider phylogenomic relationships, as well as to further explore population genomics of the system and investigate the peaks of genomic differentiation between species (Ford *et al*. in prep).

A broader phylogenetic approach may also allow dating of the soda lake colonisation. A fossil record is available for the wider genus *Oreochromis* genus

(discussed in chapter one), and *O. lorenzoi* ~6.0 Ma (Carnevale *et al.* 2003) has previously been used as a fossil constraint in molecular analyses (Schwarzer et al. 2009). A phylogeny of the genus will also help to clarify the relationship of *Alcolapia* to *Oreochromis*, as previous phylogenetic treatments have been based on mtDNA alone (e.g., Kavembe *et al.* 2013), or only included few *Oreochromis* species (e.g., Schwarzer *et al.* 2009; Dunz & Schliewen 2013). Resolving this relationship will also provide information on whether the adaptation to soda conditions in *Alcolapia* and *Oreochromis* spp. arose convergent or independently. A multilocus phylogeny of *Oreochromis* is currently in progress, with sequence data for 26 out of 33 *Oreochromis* species (Ford *et al*. in prep). The phylogeny employs molecular markers that accumulate mutations rapidly, to take account of the young age of *Oreochromis* (Schwarzer *et al.* 2009) and recent radiation of *Alcolapia* (Seegers *et al*. 2001). The multilocus dataset employs the following mtDNA markers: i) control region 1013bp – the non-coding region of mtDNA, which accumulates mutations at a faster rate than other mtDNA loci and therefore may be best suited to resolving recently separated taxa (McMillan & Palumbi 1997); ii) NADH dehydrogenase subunit 2 (ND2) 1047bp – a marker that is frequently used in cichlid phylogenetic studies, and has provided resolution between recently diverged taxa (Kocher *et al*. 1995; Nagl *et al*. 2001; Klett & Meyer 2002; Won *et al*. 2006; Schwarzer *et al*. 2009); and the following nuclear markers: bmp4 (577bp), tyr (659bp), gapdhs (499bp), ccng1 (647bp) – four loci for which primers have recently been developed for young radiations, whereby primer binding is designed for amplification of both the gene and flanking intron regions, which exhibit greater diversity than the respective exons (Meyer & Salzburger 2012); and the (non-coding region) first intron of the ribosomal protein coding gene s7 (500bp) – the first intron of s7 has proven useful in phylogenies of closely related species, and has shown the greatest resolution of several nuclear markers in various cichlid and freshwater fish phylogenies (Chow & Hazama 1998; Schwarzer *et al*. 2009; Liu *et al*. 2012; Wright *et al*. 2012).

Chapter four demonstrated the detection of outlier loci in several species comparisons, but only one of the intraspecific comparisons (*A. alcalica* northern and southern clades). Detecting outlier loci in such narrowly differentiated intraspecific comparisons is inherently difficult due to the low levels of genome-wide differentiation, but continuing developments in analysis methodologies may provide increased power to detect such discrepancies. A recent update to the BayeScan algorithm (BayeScan 3; methods published but software not yet released; Foll *et al.* 2014) allows hierarchical modelling for multiple comparisons (rather than the pairwise comparisons considered in chapter four) and may have more power to

detect outliers with a lower false discovery rate in datasets such as this (O. Gaggiotti; pers comms.). Other recent software developments include models that account for evolutionary nonindependence and those that consider correlation with environmental variables such as FLK (Bonhomme *et al.* 2010), Bayenv2 (Günther & Coop 2013), BayScEnv (de Villemereuil 2015), all of which may provide further power to detect loci under spatially divergent selection (Lotterhos & Whitlock 2014). Furthermore, analysis using multiple models to detect outlier loci have been shown to provide the lowest false discovery rates (de Villemereuil *et al.* 2014). Additionally, increased coverage of the regions of interest either through the use of whole genome resequencing or targeted sequencing (e.g., Nadeau *et al.* 2012; Martin *et al.* 2013) may improve the resolution of this differentiation. Whole genome resequencing will also increase the number of markers available for alternative testing such as genome-wide association studies. Increased marker density would be relevant here, especially as patterns of LD were high across all loci (linkage seen at distances of up to 200~500 Kb, chapter three) and LD can bias such association studies (Pardo-Diaz *et al.* 2015).

A related goal would be to identify regions of the genome that are responsible for the phenotypic diversification and polymorphism observed in these cichlid fishes, despite the very shallow divergence between species and determining whether these underlie peaks of genomic differentiation. Characterising the genomic architecture and structural variation underlying regions of differentiation (e.g., insertions, deletions, inversions) would also provide information on how these changes may have arisen in a recent diversification. Although the present study highlights heterogeneous genomic differentiation, more detailed analysis is required to identify the regions of high divergence and assess their impact on species differentiation. Determining whether regions of species differentiation have arisen de novo or are based on standing genetic variation may also be relevant here, given the recent indication of the importance of existing genetic variation and divergent selection acting on many genes in cichlid speciation (Brawand *et al*. 2014).

The present analysis is based entirely on wild-caught, preserved specimens. However, several additional questions could be addressed with the establishment of aquarium stocks. Determining the degree of reproductive isolation between the *Alcolapia* species will be critical to confirming species status, as well as the inference of adaptive radiation and ecological speciation in this system, and testing the importance of sexual selection in the system based on male colouration The degree of isolation could be tested with aquarium populations using behavioural mate choice tests (e.g. Tyers & Turner 2013; Selz *et al.* 2014) between species, and

would also allow examination of the importance of intraspecific morphological and colour variation between populations to reproductive isolation. As such, testing isolation between allopatric populations would allow inference of the mechanism of isolation, for example whether assortative mating is stronger in southern Natron *A. alcalica* populations than in northern populations where other species are not encountered. Furthermore, testing the degree of reproductive isolation between lineages at varying levels of divergence would allow inference of the rate of isolation development, and assessment of which barriers occur early and remain throughout the speciation continuum (Coyne & Orr 2004; Merrill *et al.* 2011). Assortative mating in cichlids has previously been suggested to be enacted by visual (Genner *et al.* 2007; Seehausen *et al.* 2008), olfactory (Plenderleith *et al.* 2005; Blais *et al.* 2009) and audio (Verzijden *et al.* 2010) cues, which could be further investigated here. Aquarium populations would also allow several other questions to be addressed: the relative fitness and viability of F1 hybrid crosses; the heritability of the characteristic species trophic morphologies; and the plasticity of trophic morphology in response to food and substrate type. Testing plastic response to food type would be pertinent here, as oral jaw morphology in cichlids is known to be largely controlled by only a few genes (Albertson *et al.* 2003), and adaptive divergence in jaw shape has been seen across the space of a few generations (van Rijssel *et al.* 2014). Furthermore, jaw and tooth shape are known to exhibit plastic responses to food type, and may thus quickly respond to habitat changes. Finally, phenotypic plasticity can facilitate ecological diversification, by allowing novel phenotypes to survive in novel resource environments (Pfennig *et al.* 2010; Burress 2014), and body shape has been shown to be controlled by both plastic and genetic components in ecological morphs of Tanganyika cichlids (e.g.,. Theis *et al*. 2014). Although other *Oreochromis* species exhibit thickened or enlarged lips (e.g., *O. urolepis*), this is a secondary male characteristic only displayed in breeding males, and no other *Oreochromis* species exhibit the extensive morphological variation or subterminal mouths displayed in *Alcolapia* (Trewavas 1983). Understanding the basis for such divergent morphology may provide information on whether these features were important in the colonisation of the lakes and whether they were instrumental in species divergence within the soda lakes. If the morphology is a result of differential gene expression between lineages, as has been shown to be the case in thick- and thin-lipped morphs of both African and Neotropical cichlids (Colombo et al. 2013; Manousaki et al. 2013), then transcriptomic analysis (RNA-seq) would be a useful tool to investigate the candidate genes underlying the morphogenesis of these traits.

Although reproductive isolation has classically been tested using mate choice trials, a complementary approach could be employed in this system using field observations and genotyping. Collection of mated cichlid pairs from wild populations has previously been used to infer assortative mating by body shape and diet (Martin 2013). Furthermore, collection of mouthbrooding females could be used to genotype each female's offspring and assign paternal species or test whether entire broods exhibit paternity of multiple species. Microsatellite genotyping has also been used to assign parentage in aquarium mate choice trials (e.g., Genner *et al*. 2007), and more recently low-density (~100 SNPs) SNP arrays have been used to examine parentage in livestock and aquaculture (e.g., Fernandez *et al.* 2013), as well as wild populations (Norman *et al.* 2013), and shown to be robust to unsampled parental genotypes in fish populations (Steele *et al.* 2013). Although a SNP array is not currently available for the focal species, whole genome sequence data could be used to develop an array of SNPs characteristic of and segregating each species/morph. A subsequent SNP-Chip based approach could then be used to genotype fish fry collected from brooding females, which has the benefit of not only being substantially cheaper than genotyping-by sequencing (e.g., ~£1000 to screen ~100 SNPs in 200 individuals, vs. £3000 to sequence whole genome data for four individuals), but also requires substantially less genomic DNA, which would be a consideration in the use of small tissue samples. Such a field-based approach would allow reproductive isolation to be tested in the natural environment rather than in artificial laboratory conditions, and could provide validation of aquarium mate choice trials.

Further field surveys and observations are warranted to fully describe the variation across the soda lake system, especially as some of the most isolated populations with highest levels of divergence seen in the present study (e.g., sites 15 and 19) occur in the least-surveyed areas of Lake Natron, with much of the northern area entirely unexplored (although also the most inaccessible). Determining the occurrence of *Alcolapia* populations on the western shore of Lake Natron at the point where the Peninj river enters the lake basin would allow comparison with the population at the site of the Ewaso Ngiro influx (site 19). Additional field observations at the site of proposed ecological sexual divergence (site 15) could confirm whether individuals segregate by sex in the water column, and increased sampling could corroborate the correlation of mouth morphology with sex. Additionally, environmental measurements would clarify the environmental influence on the cranial morphological differentiation relative to other sites. The morphological overlap at site 14 warrants genetic investigation to determine the

extent of hybridisation at this site. Another site identified by Seegers *et al.* (2001) (Seegers *et al.* 2001) (not visited in the present study, but south of site 14 in this study) was inhabited by individuals with intermediate morphology of *A. alcalica* and *A. ndalalani*, suggesting that populations found on the south-east coast experience hybridisation, and this area would benefit from further investigation. Additional ecological and behavioural surveys of the existing sites could also provide more information on species interactions, including: ecological segregation by examining species abundance along springs; observations and underwater filming of feeding behaviour in allopatry and sympatry to consider microhabitat differentiation; intra- and interspecific male aggression in natural conditions.

The increasing availability of genomic resources for cichlid species, and in particular the reference genome of closely related *O. niloticus* (Brawand *et al*. 2014), means that there is considerable insight to be gained from genomic approaches to study speciation in *Alcolapia*. The generation of densely sampled genomic datasets, alongside field, behavioural, and ecological studies, will be of substantial benefit to elucidating the processes initiating and maintaining speciation and adaptive divergence in this unique example of radiation in one of the Earth's most hostile environments.

References

Abbott R, Albach D, Ansell S *et al*. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.

Albertson RC, Streelman JT, Kocher TD (2003) Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 5252–5257.

Bayona J, Akinyi E (Eds.) (2006) *Alcolapia grahami. The IUCN Red List of Threatened Species.*

Bayona RJD (Ed.) (2006) *Alcolapia alcalicus. The IUCN Red List of Threatened Species.*

Blais J, Plenderleith M, Rico C *et al.* (2009) Assortative mating among Lake Malawi cichlid fish populations is not simply predictable from male nuptial colour. *BMC Evolutionary Biology*, **9**, 53.

Bonhomme M, Chevalet C, Servin B *et al.* (2010) Detecting selection in population trees: The Lewontin and Krakauer test extended. *Genetics*, **186**, 241–262.

Brawand D, Wagner CE, Li YI *et al.* (2014) The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, **513**, 375–381.

Burress ED (2014) Cichlid fishes as models of ecological diversification: patterns, mechanisms, and consequences. *Hydrobiologia*, **748**, 7–27.

Carnevale G, Sorbini C, Landini W (2003) *Oreochromis lorenzoi*, a new species of Tilapiine cichlid from the Late Miocene of Central Italy. *Journal of Vertebrate Paleontology*, **23**, 508–516.

Chow S, Hazama K (1998) Universal PCR primers for S7 ribosomal protein gene introns in fish. *Molecular Ecology*, **7**, 1255–1256.

Coe MJ (1966) The Biology of *Tilapia grahami* Boulenger in Lake Magadi, Kenya. *Acta Tropica*, **23**, 146–198.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, MA.

de Villemereuil P (2015) BayeScEnv software program. Available at: https://github.com/devillemereuil/bayescenv, Accessed July 2015.

de Villemereuil P, Frichot É, Bazin É *et al*. (2014) Genome scan methods against more complex models: when and how much should we trust them? *Molecular Ecology*, **23**, 2006–2019.

Dunz AR, Schliewen UK (2013) Molecular phylogeny and revised classification of the haplotilapiine cichlid fishes formerly referred to as “Tilapia.” *Molecular Phylogenetics and Evolution*, **68**, 64–80.

Faria R, Renaut S, Galindo J *et al.* (2014) Advances in ecological speciation: an integrative approach. *Molecular Ecology*, **23**, 513–521.

Fernandez ME, Goszczynski DE, Liron JP *et al.* (2013) Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd. *Genetics and Molecular Biology*, **36**, 185–191.

Foll M, Gaggiotti OE, Daub JT *et al*. (2014) Widespread signals of convergent adaptation to high altitude in Asia and America. *The American Journal of Human Genetics*, **95**, 394–407.

Genner MJ, Nichols P, Carvalho GR *et al.* (2007) Reproductive isolation among deep-water cichlid fishes of Lake Malawi differing in monochromatic male breeding dress. *Molecular Ecology*, **16**, 651–662.

Günther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. *Genetics*, **195**, 205–220.

Hendry AP (2009) Ecological speciation! Or the lack thereof? *Canadian Journal of Fisheries and Aquatic Sciences*, **66**, 1383–1398.

Hudson AG, Vonlanthen P, Seehausen O (2011) Rapid parallel adaptive radiations from a single hybridogenic ancestral population. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 58–66.

Kadigi RMJ, Kilima FT, Kashaigili JJ (2012) *A comparative study of costs and benefits of soda ash mining and promotion of ecotourism and sustainable use of natural resources in Lake Natron basin, Tanzania*. Birdlife International, Birdlife Africa Partnership Secretariat, Nairobi, Kenya.

Kavembe GD, Machado-Schiaffino G, Meyer A (2013) Pronounced genetic differentiation of small, isolated and fragmented tilapia populations inhabiting the Magadi Soda Lake in Kenya. *Hydrobiologia*, **739**, 55–71.

Klett V, Meyer A (2002) What, if anything, is a Tilapia? - Mitochondrial ND2 phylogeny of tilapiines and the evolution of parental care systems in the African cichlid fishes. *Molecular Biology and Evolution*, **19**, 865–883.

Kocher TD, Conroy JA, McKaye KR *et al*. (1995). Evolution of NADH dehydrogenase subunit 2 in East African cichlid fish. *Molecular Phylogenetics and Evolution*, **4**, 420–432.

Lamichhaney S, Berglund J, Almén MS *et al.* (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, **518**, 371–375.

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology*, **23**, 2178–2192.

Liu D, Guo H-Y, Tang W-Q, Yang J-Q (2012) Comparative evolution of S7 intron 1 and ribosomal internal transcribed spacer in *Coilia nasus* (Clupeiformes: Engraulidae). *International Journal of Molecular Sciences*, **13**, 3085–3100.

Mallet J (1995) A species definition for the Modern Synthesis. *Trends in Ecology & Evolution*, **10**, 294–299.

Martin CH (2013) Strong assortative mating by diet, color, size, and morphology but limited progress toward sympatric speciation in a classic example: Cameroon crater lake cichlids. *Evolution*, **67**, 2114–2123.

Martin CH, Cutler JS, Friel JP *et al.* (2015) Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. *Evolution*, doi: 10.1111–evo.12674.

Martin SH, Dasmahapatra KK, Nadeau NJ *et al.* (2013) Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, **23**, 1817–1828.

McMillan WO, Palumbi SR (1997) Rapid rate of control-region evolution in Pacific butterflyfishes (Chaetodontidae). *Journal of Molecular Evolution*, **45**, 473–484.

Merrill RM, Gompert Z, Dembeck LM *et al.* (2011) Mate preference across the speciation continuum in a clade of mimetic butterflies. *Evolution*, **65**, 1489–1500.

Meyer BS, Salzburger W (2012) A novel primer set for multilocus phylogenetic inference in East African cichlid fishes. *Molecular Ecology Resources*, **12**, 1097–1104.

Nadeau NJ, Whibley A, Jones RT *et al.* (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **367**, 343–353.

Nagl S, Tichy H, Mayer WE *et al*. (2001) Classification and phylogenetic relationships of African tilapiine fishes inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, **20**, 361–374.

Norman AJ, Street NR, Spong G (2013) De novo SNP discovery in the Scandinavian brown bear (*Ursus arctos*). *PLoS ONE*, **8**, e81012.

Pardo-Diaz C, Salazar C, Jiggins CD (2015) Towards the identification of the loci of adaptive evolution. *Methods in Ecology and Evolution*, **6**, 445–464.

Pfennig DW, Wund MA, Snell-Rood EC *et al.* (2010) Phenotypic plasticity's impacts on diversification and speciation. *Trends in Ecology & Evolution*, **25**, 459–467.

Plenderleith M, Oosterhout CV, Robinson RL, Turner GF (2005) Female preference for conspecific males based on olfactory cues in a Lake Malawi cichlid fish. *Biology Letters*, **1**, 411–414.

Schliewen UK, Kocher TD, McKaye KR *et al*. (2006) Evidence for sympatric speciation? *Nature*, **444**, E12–E13.

Schluter D (2000) *The ecology of adaptive radiation*. Oxford University Press, Oxford.

Schwarzer J, Misof B, Tautz D, Schliewen U (2009) The root of the East African cichlid radiations. *BMC Evolutionary Biology*, **9**, 186.

Seegers L, Tichy H (1999) The *Oreochromis alcalicus* flock (Teleostei: Cichlidae) from Lake Natron and Magadi, Tanzania and Kenya, with description of two new species. *Ichthyological Explorations of Freshwaters*, **10**, 97–146.

Seegers L, Sonnenberg R, Tichy H (2001) The *Alcolapia* group, a remarkable species flock from lakes Natron, Tanzania, and Magadi, Kenya: a further piece of the puzzle of cichlid evolution? *Journal of Aquariculture and Aquatic Sciences*, **9**, 335–364.

Seehausen O (2004) Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, **19**, 198–207.

Seehausen O (2015). Process and pattern in cichlid radiations - inferences for understanding unusually high rates of evolutionary diversification. *New Phytologist,* doi: 10.1111–nph.13450. doi:10.1111/nph.13450

Seehausen O, Wagner CE (2014) Speciation in freshwater fishes. *Annual Review of Ecology, Evolution, and Systematics*, **45**, 621–651.

Seehausen O, Terai Y, Magalhaes IS *et al.* (2008) Speciation through sensory drive in cichlid fish. *Nature*, **455**, 620–626.

Selz OM, Pierotti MER, Maan ME *et al.* (2014) Female preference for male color is necessary and sufficient for assortative mating in 2 cichlid sister species. *Behavioral Ecology*, **25**, 612–626.

Steele CA, Anderson EC, Ackerman MW *et al.* (2013) A validation of parentage-based tagging using hatchery steelhead in the Snake River basin. *Canadian Journal of Fisheries and Aquatic Sciences*, **70**, 1046–1054.

Theis A, Ronco F, Indermaur A *et al*. (2014) Adaptive divergence between lake and stream populations of an East African cichlid fish. *Molecular Ecology*, **23**, 5304–5322.

Thys van den Audenaerde DFE (1968) *An annotated bibliography of Tilapia (Pisces, Cichlidae)*. Royal Museum Central Africa, Tervuren.

Trewavas E (1983) *Tilapiine Fishes of the genera Sarotherodon, Oreochromis and Danakilia*. British Museum (Natural History), London.

Tyers AM, Turner GF (2013) Signal and preference divergence among populations of the non-endemic basal Lake Malawi cichlid fish *Astatotilapia calliptera* (Perciformes: Cichlidae). *Biological Journal of the Linnean Society*, **110**, 180–188.

van Rijssel JC, Hoogwater ES, Kishe-Machumu MA *et al.* (2014) Fast adaptive responses in the oral jaw of Lake Victoria cichlids. *Evolution*, **69**, 179–189.

Verzijden MN, van Heusden J, Bouton N *et al.* (2010) Sounds of male Lake Victoria cichlids vary within and between species and affect female mate preferences. *Behavioral Ecology*, **21**, 548–555.

Won Y-J, Wang Y, Sivasundar A *et al*. (2006) Nuclear gene variation and molecular dating of the cichlid species flock of Lake Malawi. *Molecular Biology and Evolution,* **23**, 828–837.

Wright, JJ, David SR, Near TJ (2012) Gene trees, species trees, and morphology converge on a similar phylogeny of living gars (Actinopterygii: Holostei: Lepisosteidae), an ancient clade of ray-finned fishes. *Molecular Phylogenetics and Evolution*, **63**, 848–856.

Zaccara S, Crosa G, Vanetti I *et al.* (2014) Genetic and morphological analyses indicate high population mixing in the endangered cichlid *Alcolapia* flock of East Africa. *Conservation Genetics*, **15**, 429–440.

Appendix 7 - All samples included per analysis

GMM: Geometric morphometrics; PHJ: Pharyngeal jaw;

RAD: Restriction site-associated DNA; SIA: Stable isotope analysis;

SL: Standard length. Gut: gut length and stomach contents.

Voucher specimens are stored in the Day Lab at UCL.

Site	Taxa	ID code	SL (mm)	RAD	SIA	GMM		Gut
						Body	PHJ	
5	*A. latilabris*	005-044-AL	49		X			
		005-045-AL	55		X			
		005-086-AL	47	X	X	X	X	X
		005-087-AL	54	X	X	X	X	X
		005-088-AL	40	X	X	X		X
		005-089-AL	56		X	X		
		005-090-AL	48		X	X	X	X
		005-091-AL	44			X		
		005-092-AL	40			X		
		005-094-AL	38		X	X		X
		005-095-AL	41			X		
		005-096-AL	40			X		
		005-097-AL	48			X	X	
		005-098-AL	44		X	X		X
		005-099-AL	36			X		
		005-100-AL	49			X		
		005-101-AL	52		X	X	X	X
		005-103-AL	55		X	X	X	X
		005-105-AL	51			X		
		005-106-AL	49			X		
		005-108-AL	35		X	X	X	
		005-109-AL	55			X	X	
		005-110-AL	44			X		
		005-111-AL	46			X	X	
		005-112-AL	62	X	X	X	X	X
		005-113-AL	42			X		
		005-114-AL	51			X		
		005-115-AL	44		X	X		
		005-116-AL	53			X		X
		005-117-AL	56		X	X		X
		005-118-AL	50			X		
		005-120-AL	44			X		
		005-121-AL	44			X		
	A. ndalalani	005-046-AN	39		X			
		005-047-AN	44	X		X		
		005-048-AN	44		X			
		005-123-AN	41		X	X	X	X
		005-124-AN	39		X	X		X
		005-125-AN	40	X	X	X	X	X
		005-126-AN	32	X	X	X		
		005-127-AN	46			X		X
		005-128-AN	36		X	X	X	X
		005-129-AN	39			X	X	X
		005-130-AN	42		X	X		
		005-131-AN	41			X		

Site	Taxa	ID code	SL (mm)	RAD	SIA	GMM		Gut
						Body	PHJ	
5	*A. ndalalani*	005-132-AN	43	X	X	X	X	X
		005-134-AN	43			X	X	X
		005-135-AN	40			X		
		005-136-AN	40			X		
		005-137-AN	42			X	X	
		005-139-AN	46		X	X	X	X
		005-140-AN	44			X		
		005-141-AN	48		X	X	X	X
		005-142-AN	42			X	X	
		005-143-AN	40			X		
		005-144-AN	42			X		
		005-145-AN	36			X		
		005-146-AN	38					X
		005-147-AN	40			X		
		005-148-AN	37			X		
		005-149-AN	35			X		
		005-150-AN	42			X		X
		005-151-AN	36			X		
		005-152-AN	39			X		
		005-154-AN	37			X		
		005-155-AN	47		X	X		
		005-157-AN	39			X		
		005-158-AN	36			X		
		005-159-AN	36			X		
	A. alcalica	005-049-AA	49		X	X		X
		005-159-AA	36		X	X	X	
		005-160-AA	48	X	X	X	X	X
		005-161-AA	52		X	X	X	X
		005-162-AA	48		X	X	X	X
		005-163-AA	50	X	X	X		X
		005-164-AA	34	X	X	X		
		005-165-AA	59	X	X	X	X	X
		005-166-AA	38		X	X	X	
		005-167-AA	52		X	X	X	X
		005-168-AA	42		X	X	X	X
		005-169-AA	57		X	X	X	X
		005-170-AA	51		X	X	X	X
6	*A. alcalica*	006-050-AA	35		X			
		006-051-AA	38		X			
		006-058-AA	39	X	X	X		
		006-059-AA	43	X	X	X		
		006-060-AA	35		X	X		
		006-061-AA	33			X		
		006-062-AA	34		X	X		
		006-063-AA	37		X	X		
		006-064-AA	37		X	X		
		006-065-AA	39		X	X		
		006-066-AA	34			X		
		006-067-AA	33		X	X		
		006-068-AA	34		X	X		
		006-069-AA	36		X	X		
		006-070-AA	31			X		
		006-071-AA	35		X	X		
		006-072-AA	31			X		
		006-073-AA	34	X	X	X		
		006-074-AA	31			X		

Site	Taxa	ID code	SL (mm)	RAD	SIA	GMM		Gut
						Body	PHJ	
6	*A. alcalica*	006-075-AA	29			X		
		006-076-AA	30	X	X	X		
		006-077-AA	25			X		
		006-078-AA	30			X		
		006-079-AA	27			X		
		006-080-AA	31			X		
		006-081-AA	26			X		
07A	*A. alcalica*	07A-052-AA	38		X			
		07A-053-AA	34		X			
		07A-171-AA	36		X	X		
		07A-173-AA	35		X	X		
		07A-174-AA	30			X		
		07A-175-AA	36		X	X		
		07A-177-AA	36		X	X		
		07A-178-AA	29			X		
		07A-179-AA	37		X	X		
		07A-180-AA	33			X		
		07A-181-AA	39		X	X		
		07A-184-AA	34			X		
		07A-185-AA	32		X	X		
		07A-186-AA	37		X	X		
07B	*A. alcalica*	07B-187-AA	49		X			
		07B-188-AA	34		X			
		07B-189-AA	40		X			
		07B-190-AA	45		X			
		07B-191-AA	36		X			
		07B-192-AA	39		X			
		07B-193-AA	38		X			
		07B-194-AA	40		X			
		07B-195-AA	35		X			
		07B-852-AA	40		X			
8	*A. alcalica*	008-198-AA	65			X		
		008-199-AA	69		X			
		008-200-AA	66		X	X		
		008-201-AA	65		X	X		
		008-204-AA	43		X	X		
		008-205-AA	82		X	X		
		008-206-AA	37		X	X		
		008-207-AA	46		X	X		
		008-209-AA	43		X	X		
		008-215-AA	41			X		
		008-216-AA	61		X	X		
		008-217-AA	40		X	X		
		008-219-AA	41			X		
		008-220-AA	45		X	X		
		008-223-AA	36			X		
		008-225-AA	47		X			
		008-227-AA	34		X	X		
		008-228-AA	30			X		
9	*A. alcalica*	009-229-AA	58		X	X		
		009-230-AA	58	X	X	X		
		009-231-AA	72	X	X			
		009-232-AA	52		X	X		
		009-233-AA	48			X		
		009-234-AA	48	X	X	X		
		009-235-AA	39		X			

Site	Taxa	ID code	SL (mm)	RAD	SIA	GMM		Gut
						Body	PHJ	
9	*A. alcalica*	009-236-AA	53		X	X		
		009-237-AA	47			X		
		009-238-AA	65		X	X		
		009-239-AA	64		X			
		009-240-AA	46			X		
		009-241-AA	50		X	X		
		009-242-AA	63		X			
		009-243-AA	62	X	X	X		
		009-244-AA	46		X	X		
		009-245-AA	39		X	X		
		009-246-AA	41		X	X		
	A. latilabris	009-056-AL	44		X			
		009-057-AL	34		X			
		009-247-AL	34	X	X	X		
		009-248-AA	38	X	X	X		
		009-249-AL	37	X	X	X		
		009-250-AL	43			X		
		009-251-AL	40		X	X		
		009-253-AL	38		X	X		
		009-255-AL	41		X	X		
		009-256-AL	39		X	X		
		009-257-AL	35			X		
		009-258-AL	42		X	X		
		009-259-AL	46		X			
		009-260-AL	45	X	X	X		
		009-263-AL	41			X		
		009-264-AL	41		X			
		009-265-AL	41		X	X		
		009-266-AL	38		X	X		
		009-267-AL	42			X		
		009-268-AL	42		X	X		
11A	*A. ndalalani*	11A-291-AN	34			X		
		11A-292-AN	38		X	X		
		11A-293-AN	40	X	X	X		
		11A-294-AN	40		X	X		
		11A-295-AN	40		X	X		
		11A-296-AN	40	X	X	X		
		11A-297-AN	37		X	X		
		11A-298-AN	37			X		
		11A-299-AN	32			X		
		11A-300-AN	32		X	X		
		11A-301-AN	37	X	X	X		
		11A-302-AN	35			X		
		11A-303-AN	38		X	X		
		11A-304-AN	40		X	X		
		11A-305-AN	35			X		
		11A-306-AN	37		X	X		
		11A-307-AN	38			X		
		11A-308-AN	31	X	X	X		
		11A-309-AN	37			X		
		11A-311-AN	32			X		
		11A-312-AN	36		X	X		
		11A-313-AN	40		X	X		
		11A-315-AN	38		X	X		
		11A-316-AN	38			X		

Site	Taxa	ID code	SL (mm)	RAD	SIA	GMM		Gut
						Body	PHJ	
11A	*A. alcalica*	11A-317-AA	39		X	X		
		11A-320-AA	90	X	X	X		
		11A-321-AA	86	X	X	X		
		11A-322-AA	90	X	X	X		
		11A-323-AA	92	X	X			
		11A-324-AA	87	X	X	X		
		11A-325-AA	87	X	X			
		11A-326-AA	84	X	X	X		
		11A-327-AA	30			X		
		11A-328-AA	35			X		
		11A-329-AA	40		X	X		
		11A-330-AA	36			X		
		11A-331-AA	36			X		
		11A-332-AA	39		X	X		
		11A-333-AA	32			X		
		11A-334-AA	31			X		
		11A-335-AA	36			X		
		11A-336-AA	32			X		
		11A-337-AA	32			X		
		11A-338-AA	37			X		
		11A-339-AA	57	X	X	X		
		11A-340-AA	53		X	X		
		11A-341-AA	42		X	X		
		11A-342-AA	34			X		
		11A-343-AA	32			X		
		11A-344-AA	34			X		
		11A-345-AA	31			X		
		11A-346-AA	31			X		
		11A-347-AA	29			X		
		11A-348-AA	30			X		
		11A-349-AA	39			X		
		11A-350-AA	40		X	X		
		11A-351-AA	36		X	X		
		11A-352-AA	37			X		
		11A-353-AA	37			X		
		11A-354-AA	33			X		
		11A-355-AA	32			X		
		11A-356-AA	32			X		
		11A-357-AA	32			X		
		11A-358-AA	31			X		
		11A-359-AA	33			X		
		11A-360-AA	31			X		
		11A-361-AA	30			X		
		11A-362-AA	31			X		
		11A-364-AA	30			X		
		11A-365-AA	27			X		
		11A-366-AA	28			X		
		11A-367-AA	25			X		
		11A-368-AA	25			X		
11B	*A. alcalica*	11B-372-AA	51			X		
		11B-373-AA	46			X		
		11B-374-AA	37			X		
		11B-379-AA	35			X		
		11B-381-AA	42			X		
		11B-382-AA	39			X		
		11B-383-AA	46			X		

Site	Taxa	ID code	SL (mm)	RAD	SIA	GMM		Gut
						Body	PHJ	
11B	*A. alcalica*	11B-384-AA	39			X		
		11B-385-AA	37			X		
		11B-387-AA	36			X		
		11B-388-AA	33			X		
		11B-389-AA	32			X		
	A. ndalalani	11B-390-AN	45			X		
		11B-391-AN	33			X		
		11B-392-AN	45			X		
		11B-393-AN	32			X		
		11B-394-AN	48			X		
		11B-395-AN	30			X		
		11B-396-AN	45			X		
		11B-397-AN	35			X		
		11B-398-AN	36			X		
		11B-399-AN	30			X		
		11B-401-AN	34			X		
		11B-402-AN	30			X		
		11B-403-AN	27			X		
		11B-404-AN	34			X		
		11B-405-AN	41			X		
		11B-406-AN	38			X		
		11B-407-AN	32			X		
	A. latilabris	11B-408-AL	58	X	X	X		
		11B-409-AL	54	X	X	X		
		11B-410-AL	63			X		
		11B-411-AL	45			X		
		11B-412-AL	55			X		
		11B-413-AL	59			X		
		11B-414-AL	46			X		
		11B-415-AL	56			X		
		11B-416-AL	55			X		
		11B-417-AL	52			X		
		11B-418-AL	45			X		
		11B-419-AL	41	X	X	X		
		11B-420-AL	38			X		
		11B-421-AL	54			X		
		11B-422-AL	48			X		
		11B-423-AL	51			X		
		11B-424-AL	53	X	X	X		
		11B-425-AL	48			X		
		11B-427-AL	44			X		
		11B-428-AL	47			X		
		11B-429-AL	34			X		
11C	*A. latilabris*	11C-430-AL	57		X	X		
		11C-431-AL	40			X		
		11C-432-AL	35			X		
		11C-433-AL	42			X		
		11C-434-AL	45		X	X		X
		11C-435-AL	25			X		
		11C-436-AL	31			X		
		11C-437-AL	28			X		
		11C-439-AL	43		X	X		X
		11C-440-AL	47		X	X		X
		11C-441-AL	41			X		
		11C-442-AL	48		X	X		X
		11C-443-AL	33			X		

Site	Taxa	ID code	SL (mm)	RAD	SIA	GMM		Gut
						Body	PHJ	
11C	*A. latilabris*	11C-444-AL	35			X		
		11C-445-AL	53		X	X		X
		11C-446-AL	49		X	X		X
		11C-447-AL	40			X		
		11C-448-AL	41			X		
		11C-449-AL	51		X	X		X
		11C-450-AL	51		X	X		X
		11C-451-AL	46		X	X		X
		11C-452-AL	40			X		
		11C-453-AL	41			X		
		11C-454-AL	48		X	X		
		11C-455-AL	39		X	X		
		11C-456-AL	32			X		
		11C-457-AL	46			X		
		11C-458-AL	33			X		
		11C-459-AL	30			X		
		11C-460-AL	49		X	X		X
		11C-461-AL	*28*			X		
		11C-462-AL	*37*			X		
		11C-463-AL	*31*			X		
		11C-464-AL	*37*			X		
		11C-465-AL	*29*			X		
		11C-466-AL	46		X	X		
		11C-467-AL	38			X		
		11C-468-AL	28			X		
		11C-469-AL	35			X		
		11C-470-AL	26			X		
		11C-471-AL	29			X		
		11C-472-AL	38			X		
		11C-473-AL	36			X		
		11C-474-AL	40			X		
		11C-475-AL	40			X		
		11C-476-AL	43		X	X		X
	A. ndalalani	11C-477-AN	34		X	X		
		11C-478-AN	30		X	X		X
		11C-479-AN	28			X		
		11C-480-AN	34		X	X		X
		11C-481-AN	30		X	X		X
		11C-482-AN	31			X		
		11C-483-AN	27			X		
		11C-484-AN	27			X		
		11C-485-AN	30			X		
		11C-488-AN	36		X	X		X
		11C-489-AN	26			X		
		11C-490-AN	33		X	X		X
		11C-491-AN	32		X	X		X
		11C-493-AN	31			X		
		11C-494-AN	30		X			
		11C-495-AN	34		X	X		X
		11C-496-AN	32		X	X		X
		11C-497-AA	35		X	X		
		11C-498-AN	33		X	X		X
		11C-499-AN	32		X	X		X
		11C-500-AN	33		X	X		X
		11C-501-AN	30			X		
		11C-502-AN	32		X			X

Site	Taxa	ID code	SL (mm)	RAD	SIA	GMM		Gut
						Body	PHJ	
11C	*AN*	11C-503-AN	27		X	X		
	A. alcalica	11C-504-AA	34		X	X		X
		11C-505-AA	27		X	X		X
		11C-506-AA	31		X	X		X
		11C-507-AA	43		X	X		X
		11C-508-AA	31		X	X		
		11C-509-AA	29		X	X		X
		11C-510-AA	27		X	X		
		11C-511-AA	36		X	X		X
		11C-512-AA	33		X	X		X
		11C-513-AA	31		X	X		X
		11C-514-AA	28		X	X		
		11C-515-AA	36		X	X		X
		11C-516-AA	39		X	X		X
012	*A. alcalica*	012-517-AA	79		X	X		X
		012-518-AA	73	X	X	X	X	X
		012-519-AA	64		X	X	X	X
		012-520-AA	69		X	X	X	X
		012-521-AA	74		X	X		X
		012-522-AA	59		X	X		X
		012-523-AA	56		X	X		X
		012-524-AA	46			X	X	X
		012-526-AA	104	X	X		X	X
		012-528-AA	62			X	X	
		012-529-AA	84		X		X	X
		012-530-AA	68			X	X	
		012-531-AA	54		X			X
		012-532-AA	95		X		X	
		012-533-AA	89		X	X	X	
		012-534-AA	50			X	X	
		012-535-AA	46		X			
		012-536-AA	45			X		
		012-537-AA	65		X	X	X	X
		012-538-AA	38			X		
		012-539-AA	52		X	X		
	A. ndalalani	012-540-AN	37			X	X	
		012-541-AN	38			X		X
		012-542-AN	47		X	X	X	X
		012-543-AN	41	X	X	X	X	
		012-544-AN	45		X	X	X	
		012-545-AN	36	X	X	X	X	
		012-546-AN	43		X	X	X	X
		012-547-AN	37			X		
		012-548-AN	43			X		
		012-549-AN	53		X	X		X
		012-550-AN	41		X	X		
		012-551-AN	40				X	
		012-552-AN	43			X		X
		012-553-AN	45		X	X		
		012-554-AN	54		X	X	X	
		012-555-AN	38		X	X		X
		012-556-AN	49		X	X	X	X
		012-557-AN	42			X		
		012-558-AN	46		X	X	X	X
		012-559-AN	37	X	X	X	X	
		012-560-AN	46		X	X		

Site	Taxa	ID code	SL (mm)	RAD	SIA	GMM		Gut
						Body	PHJ	
012	*A. ndalalani*	012-561-AN	36			X		
		012-562-AN	50		X	X		X
		012-563-AN	45				X	X
		012-565-AN	44					X
	A. latilabris	012-566-AL	52		X			
		012-567-AL	67		X	X	X	X
		012-568-AL	59	X	X	X	X	X
		012-569-AL	48		X	X	X	
		012-570-AL	33			X		
		012-571-AL	64			X		X
		012-572-AL	70	X	X	X	X	X
		012-573-AL	69		X	X	X	X
		012-574-AL	71		X	X	X	X
		012-575-AL	61		X	X	X	
		012-576-AL	56		X	X	X	X
		012-577-AL	59			X	X	X
		012-578-AL	57		X	X	X	
		012-579-AL	60		X	X	X	X
		012-580-AL	62		X		X	X
		012-581-AL	56			X		
		012-582-AL	54			X		
		012-583-AL	54	X	X	X	X	
		012-584-AL	57			X		
		012-585-AL	63		X	X	X	
		012-586-AL	58			X		
		012-587-AL	57		X	X	X	X
		012-588-AL	59			X		
		012-589-AL	58			X		
013	*A. alcalica*	013-590-AA	35			X		
		013-591-AA	45		X	X		
		013-592-AA	50		X	X		
		013-593-AA	51		X	X		
		013-594-AA	56		X	X		
		013-595-AA	82		X	X		
		013-596-AA	39		X	X		
		013-597-AA	73		X	X		
		013-598-AA	32		X	X		
		013-599-AA	44		X	X		
		013-600-AA	40		X	X		
		013-601-AA	36			X		
		013-602-AA	48		X	X		
		013-603-AA	52		X	X		
		013-604-AA	47		X	X		
		013-605-AA	59		X	X		
		013-606-AA	45			X		
		013-607-AA	41		X	X		
		013-610-AA	29			X		
	A. latilabris	013-612-AL	68		X	X		
		013-613-AL	40		X	X		
		013-614-AL	47			X		
		013-615-AL	40		X	X		
		013-616-AL	58		X	X		
		013-617-AL	52		X	X		
		013-618-AL	58		X	X		
		013-619-AL	67		X	X		
		013-620-AL	49		X	X		

Site	Taxa	ID code	SL (mm)	RAD	SIA	GMM		Gut
						Body	PHJ	
13	*A. latilabris*	013-621-AL	49		X	X		
		013-622-AL	54		X			
		013-623-AL	59		X	X		
		013-624-AL	39			X		
		013-625-AL	38			X		
		013-626-AL	45		X	X		
		013-627-AL	44		X	X		
		013-628-AL	56		X	X		
		013-629-AL	68		X	X		
		013-631-AL	32			X		
	A. ndalalani	013-608-AN	29			X		
		013-633-AN	37			X		
		013-634-AN	44			X		
		013-635-AN	38			X		
014	*A. latilabris*	014-636-AL	42			X		
		014-637-AL	42		X	X		
		014-638-AL	40		X	X		
		014-640-AL	42			X		
		014-641-AL	37		X	X		
		014-642-AL	34		X	X		
		014-643-AL	31			X		
		014-644-AL	47		X	X		
		014-645-AL	43			X		
		014-646-AL	39			X		
		014-647-AL	38		X	X		
		014-648-AL	35		X	X		
		014-650-AL	34			X		
		014-651-AL	44		X	X		
		014-652-AL	34			X		
		014-654-AL	40		X	X		
		014-655-AL	38			X		
		014-656-AL	41		X	X		
		014-657-AL	39		X	X		
		014-658-AL	32			X		
		014-659-AL	46			X		
		014-661-AL	43		X	X		
		014-662-AL	39		X	X		
		014-664-AL	49		X	X		
		014-665-AL	32			X		
		014-666-AL	33			X		
		014-667-AL	36		X	X		
		014-669-AL	34			X		
		014-670-AL	32			X		
		014-671-AL	35			X		
		014-672-AL	37		X	X		
		014-853-AL	36		X	X		
	A. alcalica	014-673-AA	53		X	X		
		014-674-AA	29			X		
		014-675-AA	49			X		
		014-676-AA	62		X	X		
		014-677-AA	56		X	X		
		014-678-AA	65		X	X		
		014-679-AA	48			X		
		014-680-AA	53			X		
		014-681-AA	55			X		
		014-682-AA	64		X	X		

Site	Taxa	ID code	SL (mm)	RAD	SIA	GMM		Gut
						Body	PHJ	
014	*A. alcalica*	014-683-AA	52		X	X		
		014-684-AA	54			X		
		014-685-AA	36		X	X		
		014-686-AA	50			X		
		014-687-AA	52		X	X		
		014-689-AA	62		X	X		
		014-690-AA	42			X		
		014-691-AA	55			X		
		014-692-AA	53			X		
		014-693-AA	35		X	X		
		014-694-AA	52			X		
		014-695-AA	55		X	X		
		014-696-AA	51			X		
		014-697-AA	43			X		
		014-699-AL	37			X		
		014-700-AA	43		X	X		
		014-701-AA	56			X		
		014-702-AA	47		X	X		
		014-703-AA	55			X		
		014-704-AA	52			X		
		014-705-AA	76		X	X		
		014-706-AA	51		X	X		
		014-707-AA	53			X		
015	*A. alcalica* terminal mouth morph	015-708-AA	60		X	X	X	X
		015-709-AA	42			X	X	
		015-710-AA	64		X	X	X	
		015-711-AA	48			X	X	
		015-712-AA	50		X	X	X	X
		015-713-AA	53		X	X	X	
		015-714-AA	52		X	X		
		015-715-AA	55			X		X
		015-716-AA	42			X		X
		015-717-AA	51			X		X
		015-718-AA	57		X	X		
		015-719-AA	42			X		X
		015-720-AA	59			X		
		015-721-AA	59		X	X	X	
		015-722-AA	56		X	X		X
		015-723-AA	55		X	X		
		015-724-AA	53		X	X		
		015-725-AA	49			X		
		015-726-AA	34			X		
		015-727-AA	56		X	X		
		015-728-AA	58			X		X
		015-729-AA	32			X		
		015-730-AA	31			X		
		015-848-AA	54	X	X	X	X	X
		015-849-AA	58	X	X	X	X	X
		015-850-AA	50	X	X	X	X	X
		015-851-AA	63	X	X	X	X	X
	A. alcalica upturned mouth	015-731-AU	40			X		X
		015-732-AU	41			X	X	X
		015-733-AU	52	X	X	X	X	X
		015-734-AU	60	X	X	X	X	X
		015-735-AU	51		X	X	X	X
		015-736-AU	45		X	X	X	X

Site	Taxa	ID code	SL (mm)	RAD	SIA	GMM		Gut
						Body	PHJ	
015	*A. alcalica* upturned mouth morph	015-737-AU	57		X	X	X	X
		015-738-AU	57		X	X	X	X
		015-739-AU	35			X	X	X
		015-740-AU	47		X	X		X
		015-741-AU	50		X	X		X
		015-742-AU	46		X	X		X
		015-743-AU	42			X		X
		015-744-AU	51		X	X		X
		015-745-AU	47			X		X
		015-746-AU	47		X	X		X
		015-747-AU	40			X		
		015-748-AU	51		X	X		X
		015-749-AU	46			X		X
		015-750-AU	43	X	X	X	X	X
		015-751-AU	47		X	X		X
		015-752-AU	43			X		X
		015-753-AU	51			X		X
		015-754-AU	42			X		X
		015-755-AU	35			X		X
		015-756-AU	43			X		X
		015-757-AU	42			X		X
		015-758-AU	46	X	X		X	X
		015-759-AU	44			X		X
		015-760-AU	40			X		X
		015-761-AU	43		X	X		X
017	*A. latilabris*	017-788-AL	41	X	X	X		
		017-789-AL	49	X	X	X		
		017-791-AL	57		X	X		
		017-792-AL	34			X		
		017-793-AL	47			X		
		017-794-AL	40			X		
		017-795-AL	48			X		
		017-796-AL	42			X		
		017-797-AL	45			X		
		017-798-AL	44		X	X		
		017-799-AL	45			X		
		017-800-AL	43		X			
		017-801-AL	42			X		
		017-802-AL	51		X			
		017-804-AL	45		X	X		
		017-805-AL	47			X		
		017-806-AL	39			X		
		017-807-AL	55		X	X		
		017-808-AL	49		X	X		
		017-809-AL	43			X		
		017-811-AL	39			X		
		017-812-AL	40		X	X		
		017-813-AL	61	X	X			
		017-814-AL	40			X		
		017-815-AL	57	X	X	X		
		017-816-AL	49		X			
		017-817-AL	43			X		
		017-818-AL	47		X	X		
		017-819-AL	41			X		
		017-821-AL	42			X		
		017-823-AL	42			X		

Site	Taxa	ID code	SL (mm)	RAD	SIA	GMM		Gut
						Body	PHJ	
017	*AL*	017-825-AL	35			X		
		017-827-AL	38			X		
		017-826-AL	42		X			
	A. ndalalani	017-828-AN	46	X	X	X		
		017-829-AN	34			X		
		017-830-AN	34	X	X	X		
		017-831-AN	44		X	X		
		017-832-AN	39		X	X		
		017-833-AN	40	X	X	X		
		017-834-AN	35		X	X		
		017-835-AN	38		X	X		
		017-836-AN	48		X	X		
		017-837-AN	34		X	X		
		017-838-AN	45		X	X		
		017-839-AN	42	X	X	X		
		017-840-AN	36		X	X		
		017-841-AN	34		X	X		
		017-842-AN	44			X		
		017-843-AN	43		X	X		
	A. aff n	017-844-AA	64		X			
		017-846-AA	46		X			
	AA	017-083-AA	82		X			
		017-847-AA	32		X			
018	*A. grahami*	018-854-AG	36		X			X
		018-856-AG	43		X	X		X
		018-858-AG	37		X	X		
		018-859-AG	37	X	X	X		X
		018-860-AG	41		X	X		X
		018-861-AG	40		X	X		X
		018-862-AG	41		X	X		X
		018-864-AG	40		X	X		X
		018-865-AG	37			X		
		018-866-AG	36			X		
		018-868-AG	70		X	X		X
		018-869-AG	71	X	X	X		X
		018-871-AG	65	X	X	X		X
		018-872-AG	68		X	X		X
		018-900-AG	49	X	X	X		X
		018-901-AG	38		X	X		
		018-902-AG	37		X	X		
		018-903-AG	43		X	X		X
		018-904-AG	30			X		
019	*A. alcalica*	019-874-AA	38		X	X		X
		019-875-AA	38		X	X		
		019-876-AA	54		X	X	X	X
		019-877-AA	57		X	X	X	X
		019-879-AA	68	X	X	X		X
		019-880-AA	58		X	X		X
		019-881-AA	49		X	X	X	X
		019-882-AA	52		X	X		X
		019-883-AA	51	X	X	X	X	X
		019-884-AA	49		X	X	X	X
		019-885-AA	43		X	X	X	X
		019-886-AA	42		X	X		
		019-887-AA	87	X	X	X	X	X
		019-888-AA	48	X	X	X	X	X

Site	Taxa	ID code	SL (mm)	RAD	SIA	GMM		Gut
						Body	PHJ	
021	*A. grahami*	021-895-AG	40		X	X		X
		021-897-AG	55		X			X
		021-898-AG	46	X	X	X	X	X
		021-899-AG	59	X	X		X	X
		021-905-AG	51	X	X		X	X
		021-906-AG	51		X	X	X	X
		021-907-AG	44			X	X	
		021-908-AG	42		X	X	X	
		021-909-AG	43		X	X	X	
		021-910-AG	48	X	X	X	X	X
		021-911-AG	39		X		X	X
		021-912-AG	37		X	X		X
		021-913-AG	42		X	X		
		021-993-AG	57		X	X		X
		021-995-AG	42		X	X		X
		021-994-AG	38		X	X		
022		022-976-AG	47		X	X		
		022-977-AG	38		X	X		
		022-978-AG	38		X	X		
		022-979-AG	31		X	X		
		022-980-AG	38		X	X		
		022-981-AG	26		X	X		
		022-982-AG	34		X	X		
		022-983-AG	36		X	X		
		022-984-AG	36		X			
		022-985-AG	34		X	X		
		022-986-AG	39		X	X		
		022-987-AG	34		X	X		
		022-988-AG	32		X	X		
		022-989-AG	32		X	X		
		022-990-AG	29		X	X		
023		023-918-AG	62		X	X		
		023-919-AG	61		X	X		
		023-920-AG	62		X	X		
		023-921-AG	53		X	X		
		023-922-AG	55		X	X		
		023-923-AG	70		X	X		
		023-924-AG	65		X	X		
		023-925-AG	61		X	X		
		023-926-AG	60		X	X		
		023-927-AG	52		X	X		
		023-928-AG	69		X	X		
		023-929-AG	58		X	X		
		023-930-AG	57		X	X		
		023-931-AG	59		X	X		
		023-932-AG	54		X	X		
		023-933-AG	51			X		
		023-934-AG	52			X		
024		024-940-AG	60		X	X		
		024-941-AG	65			X		
		024-942-AG	65		X	X		X
		024-943-AG	65			X		X
		024-944-AG	57		X	X		X
		024-945-AG	65		X	X		X
		024-946-AG	64		X	X		X
		024-947-AG	64		X	X		X

Site	Taxa	ID code	SL (mm)	RAD	SIA	GMM		Gut
						Body	PHJ	
24	*A. grahami*	024-948-AG	61			X		
		024-949-AG	64			X		
		024-950-AG	64	X	X	X		X
		024-951-AG	70		X	X		
		024-952-AG	55	X	X	X		X
		024-955-AG	60			X		
		024-956-AG	56		X	X		X
		024-957-AG	60			X		
		024-958-AG	59			X		
		024-959-AG	59			X		
		024-960-AG	64		X	X		X
		024-962-AG	48		X	X		
		024-963-AG	65			X		
		024-965-AG	59			X		
		024-966-AG	62			X		
		024-967-AG	56			X		
		024-968-AG	60		X			
		024-969-AG	64	X	X			X
		024-970-AG	63					X
		024-971-AG	61			X		
		024-972-AG	55		X			
		024-973-AG	44			X		
		024-975-AG	34			X		